

**IDENTIFICATION OF DNA SEQUENCE VARIANTS IN THE ESTROGEN  
RECEPTOR PATHWAY IN BREAST CANCER**

by

**Seyed Amir Bahreini**

BSc in Cell & Mol Biology, University of Isfahan, Iran, 2010

Submitted to the Graduate Faculty of

Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

**Seyed Amir Bahreini**

It was defended on

**April 6, 2016**

and approved by

**Dissertation Advisor:**

Steffi Oesterreich, PhD

Professor

Department of Pharmacology and Chemical Biology

School of Medicine

University of Pittsburgh

**Committee Members:**

Candace M. Kammerer, PhD

Professor

Department of Human Genetics

Graduate School of Public Health

University of Pittsburgh

Adrian V. Lee, PhD

Professor

Department of Pharmacology and Chemical Biology

School of Medicine

University of Pittsburgh

Ryan L. Minster, PhD

Professor

Department of Human Genetics

Graduate School of Public Health

University of Pittsburgh

Copyright © by Seyed Amir Bahreini

2016

**IDENTIFICATION OF DNA SEQUENCE VARIANTS IN THE ESTROGEN  
RECEPTOR PATHWAY IN BREAST CANCER**

Seyed Amir Bahreini, PhD

University of Pittsburgh, 2016

**ABSTRACT**

Breast cancer is of public health importance with an increasing incidence over the past decade. Estrogen Receptor (ER) activity is critical for promoting majority of breast cancers. Inhibiting ER is one of the most successful targeted therapies in oncology. Studies have suggested that genomic variation in ER binding sites and *ESR1* gene may be responsible for endocrine treatment response and cancer progression. We investigated the role of single nucleotide variants (SNVs) in the ER pathway in breast cancer, including clinically relevant mutations in ER gene and regulatory variants in ER binding sites. First, we developed a computational pipeline to identify SNVs in ER binding sites, using chromatin immunoprecipitation-sequencing (ChIP-seq) data from hormone responsive breast cancer cells and tumors. Analysis of ER ChIP-seq data from multiple MCF7 studies characterized a SNV within intron 2 of the *IGF1R* gene, rs62022087, predicted to increase the affinity for ER binding. By integrating 43 ER ChIP-seq data sets, multi-omics and clinical data, we identified SNVs regulating downstream target genes which may contribute to patients' survival. Second, we used sensitive detection methods to detect mutations and identified high frequencies *ESR1* mutations in primary tumors, metastatic lesions and cell-free DNA samples. This result may be due to higher sensitivity of our study in detecting mutations at very low allele frequency. Finally, we generated appropriate knock-in cell lines through CRISPR technology to study ER mutations. RNA-seq studies revealed ER

mutations are can activate estrogen regulated genes in a ligand independent manner and also may induce/repress a set of novel targets. Cell adhesion assays demonstrated mutants are less adhesive to Collagen I which may be a marker of metastasis. Taken together, our findings indicate that SNVs in ER pathway are clinically important and may predict drug response in ER+ breast cancer. From the public health perspective, screening for these impactful variants will be soon part of the genetic testing as our knowledge of genome improves. This will eventually help initiatives to reduce public health burden by choosing the right treatment for breast cancer patients in personalized manner.

## TABLE OF CONTENTS

<b>PREFACE.....</b>	<b>XV</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>XVII</b>
<b>1.0 INTRODUCTION.....</b>	<b>1</b>
<b>1.1 ESTROGEN RECEPTOR STRUCTURE AND FUNCTION .....</b>	<b>1</b>
<b>1.2 SIGNIFICANCE OF ESTROGEN RECEPTOR IN BREAST CANCER ....</b>	<b>3</b>
<b>1.3 DNA SEQUENCE VARIANTS ASSOCIATED WITH ER BINDING IN BREAST CANCER.....</b>	<b>5</b>
<b>1.4 <i>ESR1</i> GENE MUTATIONS IN PRIMARY AND METASTATIC BREAST CANCER.....</b>	<b>7</b>
<b>1.5 PUBLIC HEALTH RELEVANCE .....</b>	<b>9</b>
<b>1.6 HYPOTHESIS .....</b>	<b>11</b>
<b>2.0 IDENTIFICATION OF REGULATORY SINGLE NUCLEOTIDE VARIANTS IN ESTROGEN RECEPTOR BINDING* .....</b>	<b>13</b>
<b>2.1 INTRODUCTION .....</b>	<b>13</b>
<b>2.2 MATERIALS AND METHODS .....</b>	<b>16</b>
<b>2.2.1 Extracting genomic variants from ChIP-seq reads .....</b>	<b>16</b>
<b>2.2.2 Extracting somatic SNV in ER binding sites from WGS data .....</b>	<b>16</b>
<b>2.2.3 Identifying predicted DNA binding sites using ChIP-seq data .....</b>	<b>17</b>

2.2.4	Motif analysis and p-value scoring of the regSNVs* .....	17
2.2.5	TCGA data analysis.....	18
2.2.6	ChIP .....	18
2.2.7	Allele specific ChIP.....	19
2.2.8	RNA extraction and quantitative PCR (qPCR).....	19
2.2.9	Cloning and luciferase assay.....	19
2.3	RESULTS .....	20
2.3.1	<i>In-silico</i> Identification of regSNVs in MCF7 ER ChIP-seq data .....	20
2.3.2	ER binding is associated with an intronic regSNV in <i>IGF1R</i> .....	23
2.3.3	Discovery of RegSNVs in available breast cancer ER ChIP-seq data .....	26
2.3.4	Identification of somatic RegSNVs in WGS data of breast tumors .....	29
2.4	DISCUSSION.....	30
3.0	SENSITIVE DETECTION OF MONO- AND POLYCLONAL <i>ESR1</i> MUTATIONS IN PRIMARY TUMORS, METASTATIC LESIONS AND CELL FREE DNA OF BREAST CANCER PATIENTS* .....	34
3.1	INTRODUCTION .....	37
3.2	METHOD .....	39
3.2.1	Sample acquisition.....	39
3.2.2	DNA isolation, preparation, and quantification .....	39
3.2.3	Mutation detection by ddPCR.....	40
3.2.4	Quantitative analysis .....	42
3.3	RESULTS .....	42
3.3.1	<i>ESR1</i> mutations in primary tumors.....	42

3.3.2	<i>ESR1</i> mutations in bone metastases.....	44
3.3.3	<i>ESR1</i> mutations in brain metastases.....	45
3.3.4	<i>ESR1</i> mutations in cfDNA .....	45
3.3.5	Analysis of <i>ESR1</i> mutations in serial blood samples, and matched metastatic tumors .....	47
3.4	DISCUSSION.....	50
3.4.1	<i>ESR1</i> mutations are present at very low allele frequency in primary ER-positive breast cancer.....	50
3.4.2	<i>ESR1</i> is mutated in both brain and bone metastases .....	51
3.4.3	<i>ESR1</i> exhibits polyclonal mutations.....	52
3.4.4	Longitudinal monitoring of <i>ESR1</i> mutations in cfDNA.....	52
3.5	ACKNOWLEDGEMENT .....	53
4.0	THE BIOLOGY OF <i>ESR1</i> MUTATIONS IN METASTATIC BREAST CANCER* .....	55
4.1	INTRODUCTION .....	55
4.2	MATERIALS AND METHODS.....	57
4.2.1	Cell culture .....	57
4.2.2	Generation of <i>ESR1</i> mutant cell lines .....	57
4.2.3	ddPCR.....	58
4.2.4	Western blot .....	58
4.2.5	Transcriptional reporter activity of WT and mutant <i>ESR1</i> *.....	58
4.2.6	RNA-seq analysis .....	59
4.2.7	Growth Assay.....	60



4.2.8	Adhesion Assay .....	60
4.3	RESULTS .....	61
4.3.1	Molecular characterization of <i>ESR1</i> mutations.....	61
4.3.2	Transcriptome regulation by ER mutants .....	63
4.3.3	Gain of function activities of <i>ESR1</i> mutants in genome edited cell lines..	66
4.4	DISCUSSION.....	68
5.0	CONCLUSIONS .....	72
5.1	IDENTIFICATION AND FUNCTIONAL ASSESSMENT OF ER REGULATED SNVS IN BREAST CANCER.....	72
5.2	DETECTION OF <i>ESR1</i> MUTATIONS IN PRIMARY TUMORS, METASTATIC LESIONS AND CFDNA OF PATIENTS WITH ADVANCED BREAST CANCER.....	74
5.3	FUNCTIONAL ANALYSIS OF <i>ESR1</i> MUTATIONS IN ENDOCRINE TREATMENT RESISTANCE BREAST CANCER.....	76
APPENDIX A: SUPPLEMENTARY FIGURES.....		79
APPENDIX B: SUPPLEMENTARY TABLES .....		103
BIBLIOGRAPHY .....		145

## LIST OF TABLES

Table 1. A summary of <i>ESR1</i> mutations found in primary breast cancer.....	8
Table 2. Top 10 regulatory SNVs increasing ER binding to MCF7 genome. ....	23
Table 3. Top regulatory SNVs associated with the expression of their target genes.....	28
Table 4. Frequent SNVs in putative ER binding sites within WGS of 45 tumors.....	30
Table 5. The rates of <i>ESR1</i> mutations in primary tumors, cfDNA, brain and bone metastases from breast cancer patients. ....	44
Table 6. Clinical characteristics and endocrine treatment history in patients with confirmed <i>ESR1</i> mutant cfDNA, brain or bone metastases. ....	46
Table 7. The IC-50 for WT and mutant cells treated by different compounds.....	67
Table 8. Primer sets used for different assays.....	103
Table 9. The list of all ER ChIP-seq data sets in breast cancer .....	104
Table 10. The list of regulatory SNVs in MCF7 Cell line.....	106
Table 11. The list of regulatory SNVs in BT474 Cell line .....	110
Table 12. The list of regulatory SNVs in MDA-MB-134 Cell line .....	114
Table 13. The list of regulatory SNVs in T47D Cell line.....	118
Table 14. The list of regulatory SNVs in TAMR Cell line.....	122
Table 15. The list of regulatory SNVs in ZR75 Cell line. ....	126
Table 16. The list of regulatory SNVs in good prognosis tumors .....	129

Table 17. The list of regulatory SNVs in bad prognosis tumors .....	133
Table 18. The list of regulatory SNVs in metastatic tumors.....	137
Table 19. The allele frequency of top RegSNVs in ER binding sites with sufficient coverage. ....	140
Table 20. Sequence of targeted amplification primers. ....	141
Table 21. Sequence of ddPCR primer and probes. ....	142
Table 22. Cellularity and location of bone metastases. ....	143
Table 23. The sequence of sgRNA and oligos used to generate <i>ESR1</i> mutant cell lines .....	144
Table 24. Name of the novel mutant ER target genes shared between T47D and MCF7 .....	144

## LIST OF FIGURES

Figure 1. Schematic view of estrogen receptor structure.....	2
Figure 2. Schematic view of the location somatic mutation in ESR1 gene.....	8
Figure 3. The schematic view of the hypothesis model.....	12
Figure 4. Analysis pipeline for detecting regulatory SNVs from ChIp-seq data.....	21
Figure 5. The location of rs62022087 in genome and ER binding sites.....	24
Figure 6. IGF1R SNP can affect ER binding and result in higher gene expression. ....	26
Figure 7. The distribution of RegSNVs over the gnome across a panel of breast cancer cell lines, good and bad prognosis tumors. ....	27
Figure 8. Positive controls for mutation probes utilized in ddPCR technology. ....	41
Figure 9. <i>ESR1</i> mutation allele frequency of <i>ESR1</i> mutation-positive samples. ....	43
Figure 10. Clinical timelines and allele frequency of <i>ESR1</i> mutations in serial blood draws and matched metastatic lesions.....	48
Figure 11. Generation and molecular characteristics of <i>ESR1</i> mutations.....	62
Figure 12. Heatmap of ligand independent differentially regulated genes between WT and mutants.....	64
Figure 13. The overlap of ligand independent regulated genes between the cell lines for each mutations.....	64
Figure 14. Growth assay in WT and mutant cells.....	66

Figure 15. Cell adhesion to Collagen I and IV. ....	68
Figure 16. The UCSC genome browser view of the second intron in IGF1R gene. ....	79
Figure 17. The visualization of ChIP-seq reads from multiple cell lines over rs62022087 SNP site. ....	80
Figure 18. The pipeline used for analyzing breast tumors and calling somatic SNVs within ER binding sites. ....	81
Figure 19. Schematic view of somatic non-coding mutations found in two independent WGS data sets. ....	82
Figure 20. Visualization of ENCODE data around non-coding mutations sites. ....	83
Figure 21. Palindromic sequence surrounding the non-coding mutations in GPR126 and PLEKHS1 genes. ....	84
Figure 22. The gene expression of GPR126 and PLEKHS1 genes in WT vs Mut carriers. ....	85
Figure 23. Survival analysis of patients stratified by the expression of GSTM1. ....	86
Figure 24. Pre-amplification preserves mutant allele frequency and maintains sensitivity of <i>ESR1</i> -D538G mutation detection by ddPCR. ....	87
Figure 25. Pre-amplification preserves mutant allele frequency and maintains sensitivity of PIK3CA-E545K mutation detection by ddPCR. ....	88
Figure 26. <i>ESR1</i> Y537C/N/S and D538G mutation probes are specific to their corresponding mutations. ....	89
Figure 27. <i>ESR1</i> Y537C probe does not bind to wild-type allele, even at high concentrations of wild-type DNA. ....	90
Figure 28. Lack of cross-reactivity between D538G and Y537S probes. ....	91
Figure 29. LLoD determination of ddPCR. ....	92

Figure 30. Mutant allele frequency of PIK3CA H1047R mutation in 12 bone metastases.....	93
Figure 31. The D538G ER mutation in 3 brain mets was confirmed by Sanger sequencing. ....	94
Figure 32. <i>ESR1</i> Y537S and D538G observed in the same specimens are not mutated on the same alleles. ....	95
Figure 33. Clinical timelines and allele frequency of <i>ESR1</i> mutations in multiple blood draws and matched metastatic lesions from two patients.....	96
Figure 34. Sanger sequencing of <i>ESR1</i> mutations in T47D cells. ....	97
Figure 35. PCA analysis of 1000 top variable genes between WT and mutants.....	98
Figure 36. The overlap of ligand independent differentially expressed novel targets of ER. ....	99
Figure 37. Cell adhesion to Collagen I and IV in individual WT and mutant clones.....	100
Figure 38. Cell adhesion different ECMs in T47D pooled WT and mutant cells. ....	101
Figure 39. Network analysis of ligand independent novel regulated genes common between Y537S and D538G in each cell line.....	102

## **PREFACE**

Over the past few years, I have been so lucky to be surrounded by intelligent and lovely people. First and the foremost, I would like to thank my advisor, Dr Steffi Oesterreich. She was a great mentor throughout my studies and stood by me in the times of failure and success. I felt not only like a student of hers, but also a close friend, and for that I truly thank her. Further, I have always admired her knowledge and passion about science. Oftentimes, we had very productive discussions which were the key to my motivation and progress. I would also like to thank the members of my committee, Adrian Lee, Candace Kammerer and Ryan Minster for the guidance and insight that they provided. I should also thank the department faculty, in particular Candy Kammerer, who provided an excellent learning atmosphere where student progress is a priority.

I would like to thank the members of Oesterreich and Lee labs for their support, friendship and all the fun moments that we had together. My special thanks go to my colleagues and friends Kevin Levine, Rebecca Watters, Rekha Gyanchandani, Matt Sikora, Nolan Friedigkeit, Emily Harrington, Courtney Andersen, Ali Nagle, Tiffany Katz, Beth Knapick, Peilu Wang, Nilgun Tasdemir and Vaciry Li. I would like to highlight two intelligent and selfless people, Ryan Hartmaier and David Boone, who spent hours teaching me science through thought-provoking discussions. It would be hard for me to find another lab with such sincere and lovely people.

Most importantly, I like to thank my family for their endless love and support. Although it has been tough being away from them for the last few years, I would not have been able to make it through without thinking of my family. My dad, Agha Seyed, taught me how to be hard-working and take risks to achieve the best in my life. I learned how to be passionate and love other people from my mom, Shahnaz, who devoted best moments of her life taking care of me. And I have the most cheerful brother on earth, Hamid, who makes me laugh about anything. My family has lightened my way to where I stand today. I owe them what I have until the last day of my life.



## LIST OF ABBREVIATIONS

4OHT	4-hydroxytamoxifen
AAV	adeno-associated virus
AF-1	activation function 1
AF-2	activation function 2
AI	aromatase inhibitor
cfDNA	circulating free DNA
ChIP-seq	chromatin immunoprecipitation-sequencing
DBD	DNA-binding domain
ddPCR	digital droplet PCR
DE	differentially expressed
E2	17 $\beta$ -estradiol
ECM	extra cellular matrix
ER	estrogen receptor alpha
ERE	estrogen response element
GWAS	genome wide association study
LBD	ligand binding domain
LLOD	lower limits of detection

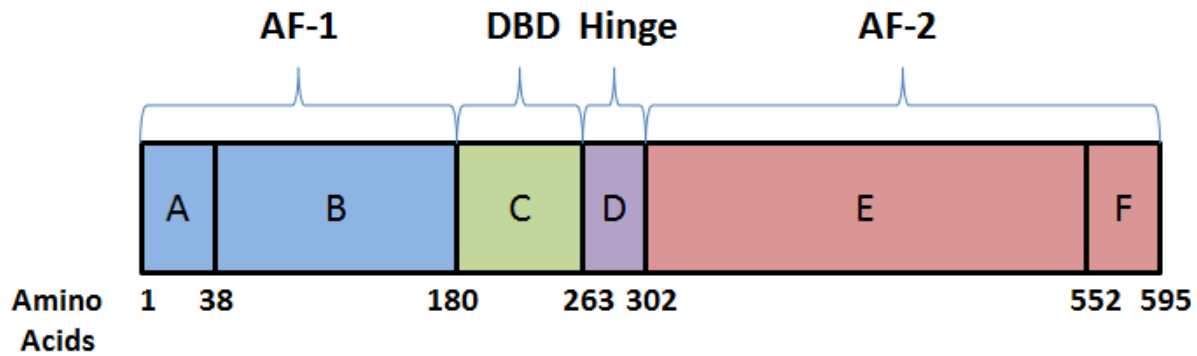
MACS	Model-based Analysis of ChIP-Seq
MAF	minor allele frequency
MAPK	mitogen-activated protein kinase
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
MPS	massively parallel sequencing
PARP	Poly(ADP-ribose) polymerase
PDX	patient derived xenograft
PGRR	Pittsburgh Genome Resource Repository
PR	progesterone receptor
qPCR	quantitative PCR
regSNP	regulatory SNP
regSNV	regulatory single nucleotide variant
SERD	selective estrogen receptor degrader
SERM	selective estrogen receptor modulator
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SRA	Sequence read archive
TCGA	The Cancer Genome Atlas
TF	transcription factor
WGS	whole genome sequencing
WT	wild-type

## **1.0 INTRODUCTION**

### **1.1 ESTROGEN RECEPTOR STRUCTURE AND FUNCTION**

Estrogen receptors are a subgroup of nuclear receptors which are responsible for sensing steroid hormones and control development, metabolism, and homeostasis of the organism (1). Two classes of estrogen receptors exist, alpha (hereinafter referred to as ER) and beta, which map to chromosomes 6q and 14q and are encoded by separate genes ESR1 and ESR2, respectively. Each receptor has distinct patterns of expression and function in normal and disease states.

ER is a nuclear transcription factor comprised of different functional domains (Figure 1). The A/B domain contains the transcriptional activation function 1 (AF-1) which is located at the NH2 terminus of the receptor and can be activated in a hormone independent manner. The C domain encompasses the DNA-binding domain (DBD), responsible for DNA interaction. The nuclear localization sequence is located in the D-domain or hinge region. Finally, the E/F domains reside in activation function 2 (AF-2) or ligand binding domain (LBD) which stimulates the receptor upon ligand binding.



**Figure 1. Schematic view of estrogen receptor structure.**

ER domains are named A-F with amino acid numbers below functional domains labeled above.

As a transcription factor, ER actively interacts with the genome to transcribe a group of target genes. There are two distinct genomic pathways that ER can exert its transcriptional activity within the cell. In the classic estrogen signaling, ER is activated by estrogen and then binds directly to estrogen response elements (EREs) to initiate the gene transcription by serving as an enhancer or repressor. Several studies have sequenced chromatin immunoprecipitation enriched DNA fragments (ChIP-seq) and have shown that ER binds to thousands of regions in the genome (2-5). These studies have revealed that the regulation of ER target genes can be mediated via proximal promoter binding or long-range interactions. In the second non-classical pathway, ER is activated by receptor tyrosine kinases which recruit transcriptional complex to the promoter of the target genes by interacting with other nuclear proteins (6). For example, ER binding sites are enriched for a number of putative binding motifs of nuclear proteins including SP-1, AP-1, Oct and C/EBP (7-9). Further, there is evidence that suggests some transcription factors can serve as a pioneer for ER binding such as FOXA1 (4) and some can control and reprogram ER chromatin binding such as progesterone receptor (PR) (10).

ER also has non-genomic mediated signaling that involves cytoplasmic proteins. Some studies have suggested the presence of ER outside the nucleus facilitating membrane and cytoplasmic signals (11). Both full length ER and alternatively spliced ER have been implicated

in cytoplasmic signaling (12,13). This non-genomic signaling results in activation of growth factor receptors, cellular tyrosine kinases, mitogen-activated protein kinases (MAPKs), phosphatidylinositol 3 kinase, and Akt (protein kinase B) -signaling enzymes, and adaptors such as adenylyl cyclases and Shc (14,15). Activation of these pathways by estrogen sends strong cell survival and cell proliferative signals via Akt and MAPK pathways. In addition, these kinases can phosphorylate ER and its coregulators to augment nuclear ER signaling (14).

## **1.2 SIGNIFICANCE OF ESTROGEN RECEPTOR IN BREAST CANCER**

Given a widespread role for estrogen signaling in human physiology, estrogen receptors have been shown to be associated with many types of abnormalities including neurodegenerative diseases, cardiovascular disorders, obesity, lupus erythematosus and several types of cancer, in particular breast cancer (16).

Breast cancer has become a major public health issue with an increasing incidence over the past decade in the US ([www.cancer.gov](http://www.cancer.gov)). Breast tumors can be classified into subtypes based on gene expression patterns among which the ER overexpressing subtypes (ER+), Luminal A and B, comprise 70-80% of all breast cancers (17-20). Interfering with estrogen action is one of the best treatments in this subset of patients with ER+ tumors. Endocrine therapy targets estrogen signaling by inhibiting ER activity or blocking the synthesis of estrogen. The selective estrogen receptor modulators (SERMs) such as Tamoxifen have been one of the major therapeutic approaches against estrogen receptor for the last 25 years (21). Tamoxifen has been shown to improve the survival as well as the quality of life in patients with breast cancer (22,23). The ER+ breast cancer became further treatable with the introduction of additional endocrine therapies

which either cut off the source of the estrogen ligand, aromatase inhibitors (AIs) such as Anastrozole and Letrozole, or degrade the ER such as Fulvestrant (24,25). Clinical trials have suggested that Letrozole and Anastrozole may be superior to Tamoxifen as the first-line therapy in hormone-receptor positive breast cancer in postmenopausal women (26-29). Fulvestrant has been also shown to be as effective as AIs; however in combination with Anastrozole, it works better in patients with metastatic disease (30-32).

Despite great advances in the treatment on ER+ breast cancer, a portion of tumors does not respond to endocrine therapy and the tumor regrows rapidly (de novo resistance). Furthermore, a substantial number of patients who do respond very well for a few years, will develop disease progression while on therapy, or even recur many years later (acquired resistance) (21,33).

Resistance to endocrine therapy is a major health and societal problem. Several studies have suggested that resistance against anti-estrogen treatment is due to crosstalk between ER and other growth factor pathways. Osborne *et al* indicated that overexpression of AIB1 and HER-2 in breast tumors is associated with worse outcome in patients undergoing tamoxifen therapy (34). Activation of the mTOR pathway may also be an alternative pathway through which tumor cells escape the effect of Tamoxifen (35). IGF1R and EGFR/MAPK pathways which are involved in cell growth and proliferation have shown to be activated in Tamoxifen resistant cell lines (36-39).

Many studies have suggested that development of resistance is caused by the tumor acquiring somatic mutations, and there is increasing evidence for a role of germline mutations. For instance, somatic mutations in PIK3CA, PTEN and TP53 and germline mutations in CYP family gene are known to be associated with endocrine response in patients (40-43).

Overall, the data suggests that a combination of genetic and transcriptomic changes could modify the response to endocrine treatments although more studies are warranted to understand the biology of *de novo* and acquired resistance in breast tumors.

### **1.3 DNA SEQUENCE VARIANTS ASSOCIATED WITH ER BINDING IN BREAST CANCER**

As mentioned previously, one of the major ways through which estrogen signaling is mediated is via ER-DNA interaction. Upon recruitment to DNA, ER facilitates the transcription or repression of downstream target genes essential for cell growth and proliferation. Our understanding of ER binding sites has been greatly improved owing to a large number of ER ChIP-seq studies in breast cancer models (2,4,5,44-46). Of importance, ER is differentially bound to DNA in Tamoxifen responsive versus resistant cell lines and tumors (4). It has also been shown that differential binding sites in breast tumors are linked to clinical outcomes in patients (4). However, the potential genomic changes underlying unique ER binding sites in different models are still unclear. A number of studies have suggested that several single nucleotide polymorphisms (SNPs) associated with breast cancer are likely to lie within EREs in the promoter of critical growth factor and cell adhesion genes such as FGFR2 and NRCAM (47-49). A computational study predicted that an ERE associated germline SNP in intron 2 of the NRCAM gene is likely to be enriched in breast cancer patients (47). Two large genome wide studies have identified risk associated SNPs which are able to generate putative ER binding site in the intron of the FGFR2 gene (48,49). In an *in silico* study, it has been indicated that breast cancer-risk associated SNPs are enriched in the binding sites of ER in a cell-type specific

manner. Testing the statistically significant SNPs in ER cistromes, the authors found a variant suppressing the expression of a downstream gene, TXO3, through modulating the FOXA1 binding to DNA (50). Clinical studies have also shown that functional SNPs in putative EREs can alter endocrine response to anti-estrogen drugs. A genome wide association study (GWAS) of breast cancer patients receiving endocrine therapy identified a SNP in the second intron of ZNF423 gene recruiting ER in the presence of 4-hydroxytamoxifen (51). A functional SNP was identified that created an ERE affecting TCL1A gene expression in a phase III trial comparing Anastrozole vs Exemstane (52). These data suggest a potential role for genomic variation underlying unique ER binding which could potentially affect the disease progression and response to anti-estrogen drugs.

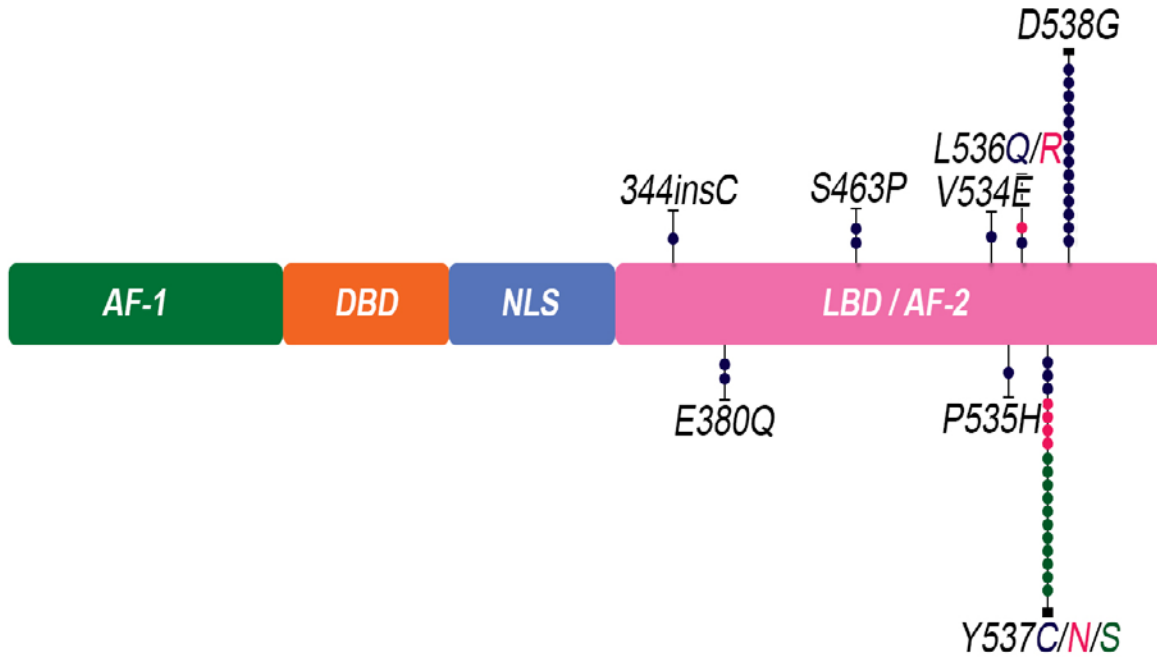
In order to better understand how SNV could alter ER binding to DNA, I developed a pipeline for identifying SNVs in ER binding sites and predicting their impact on ERE motifs. Lucas dos Santos (Department of Biomedical Informatics) helped with implementing the motif analysis in our pipeline. I used the well-established breast cancer cell line, MCF7, ER ChIP-seq data as a training sample to build our pipeline and then applied that to all available ER ChIP-seq data sets originating from breast cancer cell lines and tumors. In the following chapters, I will discuss how I employed my method to identify functionally relevant SNVs in ER bindings sites and how I validated a candidate intronic SNP in IGF1R gene by *in vitro* studies.



## 1.4 *ESR1* GENE MUTATIONS IN PRIMARY AND METASTATIC BREAST CANCER

The theory of “somatic *ESR1* mutations as a mechanism for escaping endocrine therapy” was first proposed decades ago when multiple groups tried to screen primary tumors. Few studies were able to identify *ESR1* mutations at very low frequency (53-57). Table 1 shows a summary of *ESR1* mutations found in primary breast tumors, most of which lead to a nonfunctional truncated receptor. The Cancer Genome Atlas (TCGA) also shows a mutation frequency of 1% at *ESR1* locus in about 1,000 breast tumors (58). However, there has been an exponential increase in the number of studies in the last two years describing ER as being highly mutated in metastatic breast cancer. Li *et al* first reported the incidence of *ESR1* ligand binding domain mutations in metastatic lesions (59). This was followed by a number of groups discovering a significantly higher rate of *ESR1* mutations (15-50%) in metastatic lesions derived from primary ER+ tumors (60-63). Furthermore, we and others indicated that *ESR1* mutations are detectable in the blood of patient with progressed disease (64-69). The majority of the point mutations identified in metastatic lesions are located in the LBD suggesting a gain of function for resistance to anti estrogen treatments (Figure 2). Preliminary functional studies have shown that ER mutants are hyperactive in the absence of estrogen and strongly interact with cofactors. Although it was first believed that the ER mutations arose under estrogen deprivation setting such as AI therapy, several groups found that cells transfected with *ESR1* mutant plasmids are partially resistant to tamoxifen and fulvestrant. However, both drugs showed potency to knock down ER activity at higher doses (60-63). Additionally, it has been shown that Y537S and D538G ER mutants have more affinity to ER cofactors SRC-1 and AIB1, respectively(63,70,71).

Further studies are required to better characterize the frequency of ER mutations, and their altered activities and function, in particular in the context of endocrine therapy.



**Figure 2. Schematic view of the location somatic mutation in ESR1 gene.**

The majority of mutations lie within LBD domain of ER protein with 537 and 538 amino acid sites being the most frequently mutated loci.

**Table 1. A summary of *ESR1* mutations found in primary breast cancer**

Mutation	Domain	Impact	References
N69K	AF-1	NA	(56)
A86V	AF-1	NA	(53) (53)
G160C	AF-1	NA	(72,73)
L296P	AF-2a, hinge	NA	(55)
K303R	AF-2a, hinge	Increased estrogen mediated transactivation	(74)
E352V	Hormone binding, AF-2	NA	(54) (53)
M396V	Hormone binding, AF-2	NA	(56)
437stop	HBD	NA	(54)

Based on these data, I sought to address a few urgent questions in the field:

- 1) Do rare *ESR1* mutant clones exist in primary tumors?
- 2) Are *ESR1* mutations present in understudied metastatic lesions such as bone and brain, and also circulating free DNA (cfDNA) of patients with advanced disease?
- 3) What is the biological gain of function of *ESR1* mutants in the context of hormonal resistant breast cancer?

In chapters 2 I will address questions 1 and 2 using the highly sensitive digital droplet PCR (ddPCR) methodology which is able to detect rare mutations at a frequency as low as 0.05%. This method helped us to identify a high frequency of *ESR1* mutations in primary tumors, metastatic lesions and blood of the patients with progressed disease. Finally in chapter 3, I will discuss our findings about the biology of ER mutants via integrating genomic, transcriptomic and epigenomic data.

## **1.5 PUBLIC HEALTH RELEVANCE**

Breast cancer is the most common cancer and the second most common cause of cancer death among U.S. women. Despite major advances in the early detection, diagnosis, and treatment of breast cancer, health care providers face critical challenges to create and support health care programs that can improve breast cancer outcomes. Compared to low and medium level countries, governments with well-funded health care systems have higher rates of breast cancer incidence, but also have better overall rates of breast cancer survival (75). Thus, genetic testing in breast cancer, whether for early detection or improved outcome, seems to play an important role in helping public health in these countries (76).

BRCA1 and BRCA2 mutations are probably one of the significant examples of breast cancer screening. The identification of families at the highest hereditary risk for cancer has served as a model to test strategies for prevention or early detection of breast malignancies (77,78). Genetic testing has also helped treating breast cancer patients by personalized medicine approach. For example, it has been shown that tumors deficient in BRCA and Fanconi anemia genes are more sensitive to interstrand cross-link-generating drugs (e.g., mitomycin C, platinum and its analogues) and Poly(ADP-ribose) polymerase (PARP) inhibitors (79). There are many companies offering targeted sequencing services for hotspot genes in breast cancer although the mutations from only a few genes now have prognostic value for the patients.

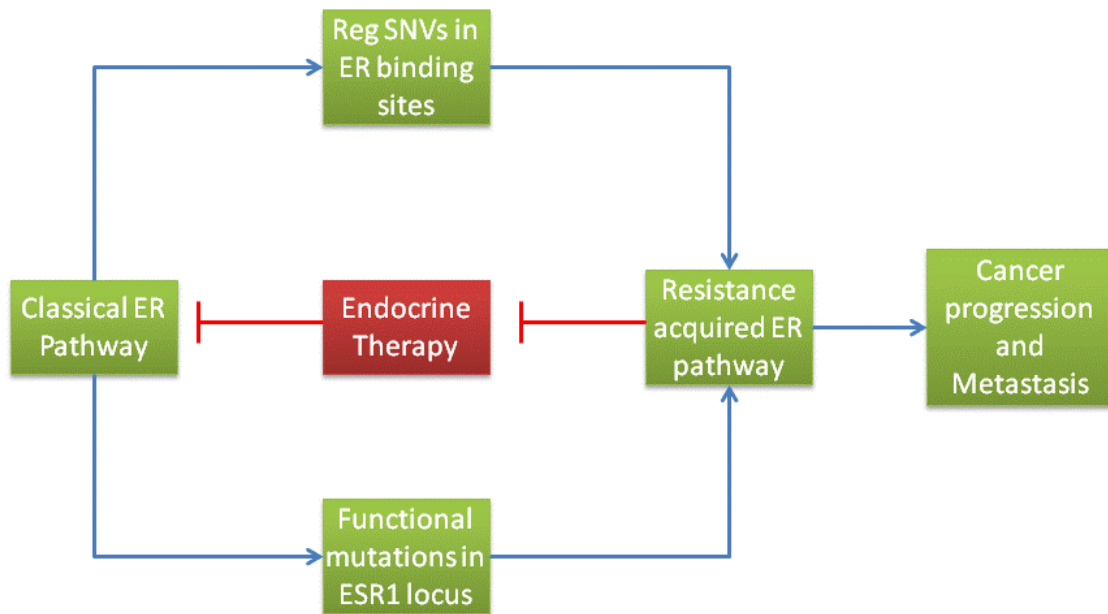
Approximately all of the commercially available genetic tests are focused on the coding genome given the poverty of our knowledge about the non-coding genome which comprises 98% of our DNA. There have been tremendous efforts in the recent years in characterizing the role of non-coding genome by large multi-center collaborative projects such as ENCODE (80). TCGA has also given us the opportunity to study non-coding DNA in cancer by sequencing hundreds of whole genomes from different tumor type (58). Two comprehensive studies of non-coding mutations in cancer demonstrated recurrent mutations in the regulatory regions of the genome (e.g. promoters, 3'UTRs and 5'UTRs) (81,82). They also indicated that these regulatory mutations could have clinical importance by impacting the survival of the patients carrying the mutations.

In breast cancer patients, about two-thirds of the tumors are ER-positive which makes them an enormous population of candidates sensitive to endocrine therapy. Our main goal in this study is to identify recurrent SNVs associated with ER pathway in breast cancer, from regulatory variants in ER binding sites to recurrent coding mutations in *ESR1* gene. Detection and tracking

the *ESR1* mutations in the blood of breast cancer patients could serve as a tool for monitoring response to endocrine therapy as testing metastatic lesions is a hassle in the clinic and could be complicated for the patients. We hope to implement the novel findings of our study to improve the predictive genetic testing in ER+ breast cancer patients in terms of therapy and survival. This will help public policy makers to direct treatment budgets more efficiently in order to target potential candidates in the setting of personalized medicine.

## **1.6 HYPOTHESIS**

I hypothesize that SNVs in ER pathway are associated with breast cancer progression and metastasis in ER+ disease (Figure 3). In my experimental model, the SNVs include *ESR1* gene mutations and genetic polymorphisms in ER binding sites. I believe such SNVs will affect ER-cofactors interaction, ER binding sites, the expression of downstream targets and eventually signaling pathways activated by ER. Our better understanding of ER associated SNVs will inform us about the biology existing behind resistance against endocrine therapy.



**Figure 3. The schematic view of the hypothesis model.**

Classical ER pathway in breast cancer is inhibited by available endocrine treatments. Two major pathways may contribute to resistance against anti E2/ER treatments and cancer progression: 1. RegSNVs in RE binding sties may modulate the ER affinity to DNA and change the expression of downstream target genes 2. Somatic mutations in ESR1 locus may alter the function of ER leading to a decreased response to endocrine therapy.

## **2.0 IDENTIFICATION OF REGULATORY SINGLE NUCLEOTIDE VARIANTS IN ESTROGEN RECEPTOR BINDING\***

**\* Dr Takis Benos and Lucas dos Santos contributed to motif analysis part in our analysis pipeline. Kevin Levine helped with analyzing TCGA data.**

### **2.1 INTRODUCTION**

Breast cancer is a major public health issue with an increasing incidence over the past decade in the US. Endocrine therapy, such as the antiestrogen tamoxifen and aromatase inhibitors, are the most successful treatment for breast cancer in which estrogen signaling is active. Estrogen signaling is mediated through ER, which upon binding the ligand estradiol, is recruited to DNA at EREs, and alters transcription of downstream target genes essential for cell growth and proliferation. The development of chromatin immunoprecipitation assays has allowed a genome-wide analysis of ER binding sites. For example, ER binds different sites in tamoxifen responsive versus resistant cell lines and tumors (4). However, the potential genomic changes underlying unique ER binding sites in different models are still unclear.

A number of studies indicate that SNPs associated with breast cancer lie within EREs, such as those in FGFR2 and NRCAM (47-49). In an *in silico* study, breast cancer-risk associated SNPs were enriched in ER binding sites in a cell-type specific manner (50). After analyzing

these statistically significant SNPs in ER cisomes, the authors found a variant suppressing the expression of a downstream gene, *TXO3*, via modulation of *FOXA1* binding to DNA (50). Clinical studies have also shown that functional SNPs in putative EREs can alter endocrine response to anti-estrogen drugs. A GWAS of breast cancer patients in a phase III trial comparing anastrozole vs exemestane identified a SNP in the second intron of *ZNF423* that is associated with recruitment of ER in the presence of 4-hydroxytamoxifen (51). A functional SNP was also identified which created an ERE conferring estrogen induction of *TCL1A* gene expression (52). These data suggest a role for genomic variation underlying unique ER binding which may affect disease progression and response to anti-estrogen therapy.

ChIP followed by high-throughput sequencing is a powerful technique for genome-wide mapping of protein-DNA interactions (83). Owing to the tremendous technological developments and reduction in the costs of the massively parallel sequencing (MPS), the number of ChIP-seq studies has grown rapidly. ChIP-seq is generally utilized to characterize the binding sites of a specific protein through enrichment of the sequencing reads over the genome. Sequencing reads have generally been simply used to identify binding sites and the strength of binding; however, recent studies have examined the actual sequences themselves, to identify variants that affect DNA binding. BCRANK is an algorithm designed to detect regulatory SNPs (regSNPs) in ChIP-chip data based upon SNP genotyping in DNA binding sites (84). More recently, another strategy used ChIP-seq data to nominate regSNPs using the assumption that the enrichment of SNPs within transcription factor (TF) binding loci indicates their regulatory function (85). This approach was applied to ENCODE data resulting in the characterization of a panel of SNPs associated with a number of transcription factors. Both approaches are primarily focused on annotated SNPs and rare variants may be missed. These studies also lack a



connection between regSNPs and the expression of *cis* target genes, which eventually determine the phenotypic output. Furthermore, appropriate motif detection could fine-tune the detection of biologically relevant variants in genome-wide binding sites.

Here we describe a pipeline integrating computational and experimental strategies to detect and validate regulatory single nucleotide variants (regSNVs) defined as germ-line or somatic single base pair changes that can affect TF binding to DNA. Our pipeline interrogates ChIP-seq reads and nominates regSNVs affecting transcription factor binding motifs. Using MCF7 cell line as the most studied model in breast cancer, we addressed whether ER binding is associated with regSNVs and resulting in differential expression of downstream genes. We further applied our computational pipeline to all available ER ChIP-seq data including ER-positive cell lines and tumors. Lastly, we modified our pipeline to accommodate discovery of somatic RegSNVs in whole genome sequencing (WGS) data. We believe that our strategy is able to identify genomic variation localized in TF binding sites having potential phenotypic significance.

## **2.2 MATERIALS AND METHODS**

### **2.2.1 Extracting genomic variants from ChIP-seq reads**

SNVs were identified from ChIP-seq data using the GATK pipeline(86). Briefly, BWA was first employed to align the raw sequence reads to the human genome reference (hg18) (87). To increase the sequence read coverage over the binding regions for more accurate variant calling, we pooled the reads from all the data sets on the same cell line. The reads were sorted and duplicates were removed using PICARD tools ([www.github.com/broadinstitute/picard](http://www.github.com/broadinstitute/picard)). To refine the mapping quality, reads were locally realigned around the known indels and finally base calls were recalibrated using GATK tools. The SNVs were called by the GATK UnifiedGenotyper tool and known variants were annotated using dbSNP and 1000genome databases. We filtered out sequence calls with a coverage <10 reads and/or a phred-score <Q20, and SNVs which were not within binding sites.

### **2.2.2 Extracting somatic SNV in ER binding sites from WGS data**

We first pooled and combined all the ER binding sites from available ChIP-seq data. This led to a comprehensive list of 331,021 binding peaks with an average length of 573 bp. Sequence read archive (SRA) files were then obtained from a WGS of 46 breast tumors paired with normal blood (88) via dbGaP(phs000472.v1.p1). BAM files were extracted from SRA files containing only the reads overlapping with ER binding sites. The generation of one bam file failed due to

quality issues of the SRA file (n=45). The trimmed bam files were then passed through GATK pipeline to mark duplicates, locally realign the reads around the indels and recalibrate the base quality score. Recalibrated bam files were used to call somatic SNVs by SomaticSniper, a package aimed to detect point mutations by comparing tumor and normal pairs (89). We finally subtracted the dbSNP variants from the Somatic calls to exclude the germline SNPs.

### **2.2.3 Identifying predicted DNA binding sites using ChIP-seq data**

The Model-based Analysis of ChIP-Seq (MACS) (90) was used to analyze all ER ChIP-seq data in breast cancer prior to July 2014 (Table 9). MACS models the length of ChIP sequencing reads to improve the resolution of predicted binding sites. We used  $1e-5$  as the p-value cutoff and assigned a genome size which matches UCSC human hg18 assembly. In data sets which had sequenced untreated genomic DNA as a control we used this sequence as input (untreated) control.

### **2.2.4 Motif analysis and p-value scoring of the regSNVs\***

For each identified SNV, sequences containing reference allele and alternative allele were computationally created. Each sequence was independently scanned using the *ESRI* human position specific matrices (PWM). The PWM was obtained from the JASPAR database (91). Determination of the potential effect of a given SNV in a binding site was inferred using reimplementations of the is-rSNP algorithm (92). Briefly, the is-rSNP calculates the background distribution of PWM scores, for a given PWM. Sequences containing reference and mutated alleles are scored and a p-value for each score is calculated. The ratio of reference and mutated

sequence p-values are calculated and compared to the background distribution of p-value ratios. If the p-value obtained from the background distribution is less than 0.05, then a SNV is considered to affect a binding site. The SNVs are next ranked based on the adjusted p-value ratio, which shows the significance of motif binding change after the introduction of the variant allele in the consensus sequence.

\*(In collaboration with Dr Takis Benos and Lucas Santana)

### **2.2.5 TCGA data analysis**

Using the Pittsburgh Genome Resource Repository (PGRR), we accessed gene expression data for 1,095 breast cancer samples and SNP array data for 501 cases. The expression of regSNVs target genes was compared between wild-type (WT) and variant carriers by a multiple comparison test. The ER positive disease was defined by ER staining in tumor samples. For the enrichment analysis, the closest adjacent genes to regSNVs were called and used to test for differential expression between ER+ and ER- tumors.

### **2.2.6 ChIP**

ChIP experiments were performed as previously described by our group (93). Briefly, hormone deprived cells were treated with 10nM E2 or vehicle (EtOH) for 45 minutes. We used ER $\alpha$  (HC-20) and rabbit IgG (sc2027) antibodies (Santa Cruz Biotechnologies) for immunoprecipitation. ChIP DNA was analyzed by qPCR using primers amplifying the rs62022087 locus in *IGF1R* (Table 8).

### **2.2.7 Allele specific ChIP**

ChIP DNA was first amplified by primers specific to the SNV site (Table 8). PCR products were TA-cloned into pCR<sup>TM</sup>4-TOPO® (Invitrogen) and plasmid was transformed to competent cells according to the manufacturer's instructions. 30 bacterial colonies were picked, DNA isolated, and subjected to Sanger sequencing. The wildtype and variant alleles were counted and the statistical significance of allele enrichment was determined by Chi-square test.

### **2.2.8 RNA extraction and quantitative PCR (qPCR)**

RNA was extracted using Illustra RNAspin Mini kit (GE Health). iScript master mix (Bio-Rad) for cDNA conversion and qPCR reactions were set up on a CFX384 thermocycler (Bio-Rad), at an annealing temperature of 60 for 40 cycles.

### **2.2.9 Cloning and luciferase assay**

ER binding sites with *IGF1R* SNP and WT alleles were amplified from MCF7 DNA using primers containing the restriction sites for EcoRV and HindIII (Table 8). PCR products and backbone plasmid pGL4-TATA-luc (pGL4.23 from Promega) were digested and ligated using thermos scientific rapid DNA ligation kit, and transformation using TOP10 competent cells. The plasmids were isolated using QIAprep Spin Miniprep Kit and further validated by Sanger sequencing.

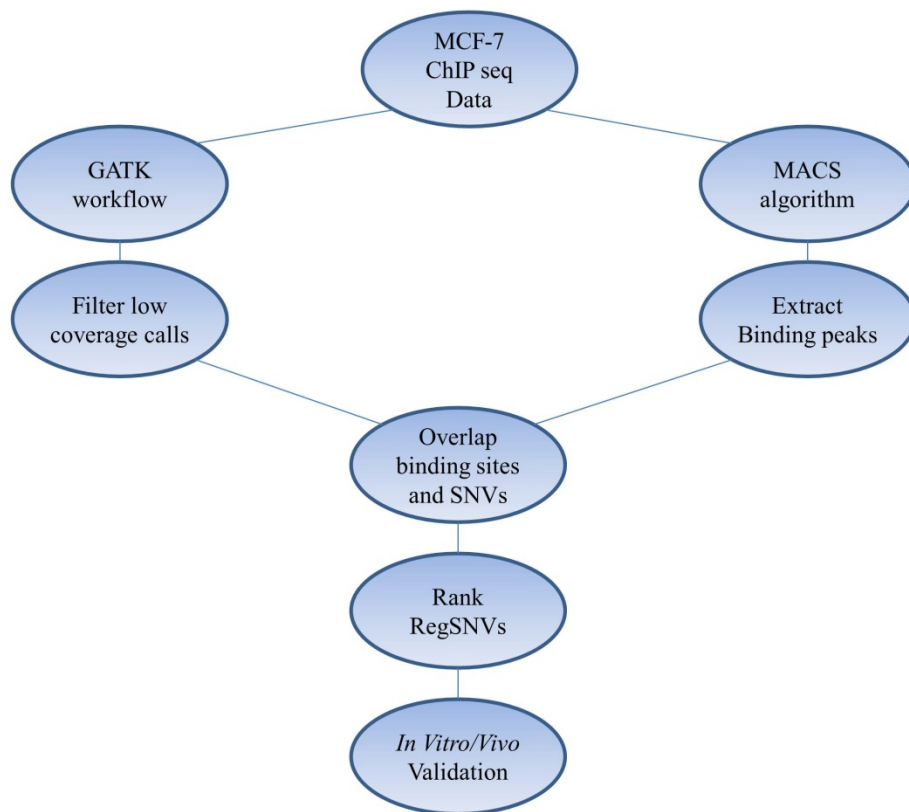
MCF7 cells were grown in DMEM, supplemented with 10% FBS. Before transfection, cells were estrogen deprived for three days with IMEM containing 10% charcoal-stripped FBS.

Cells were transfected with pGL4 ER binding site TATA –luc containing WT or SNP allele and renilla using Lipofectamine LTX with Plus. 10nM Estradiol was added to media 24hrs after transfection. Firefly and renilla luciferases were measured sequentially using the Dual-Luciferase Reporter Assay System (Promega).

## **2.3 RESULTS**

### **2.3.1 *In-silico* Identification of regSNVs in MCF7 ER ChIP-seq data**

MCF7 is one of the most employed cell lines for studying molecular genetics of breast cancer. Therefore, we selected publicly available ER ChIP-seq data from MCF7 to identify regSNVs in ER binding sites. Our computational approach (Figure 4) consisted of 1) identify SNVs from MCF7 ER ChIP-seq data, 2) identify ER binding sites using MACS, 3) overlap SNVs with ER binding sites, and 4) rank regSNVs based upon the predicted alteration of motif binding.



**Figure 4. Analysis pipeline for detecting regulatory SNVs from ChIP-seq data.**

The above pipeline was utilized to extract and rank RegSNVs based on their impact on the corresponding TF binding. MACS and GATK tools were recruited to identify binding sites and SNVs, respectively. The SNVs and binding peaks were overlapped and then regSNVs were ranked depending on how they alter EREs. One of the top candidates was selected for further functional studies.

We applied our computational workflow to nine ER ChIP-seq data sets from five different studies of MCF7 cells performed under similar experimental conditions (Table 9) (3-5,45,46). 303,964,039 reads were mapped to the human genome (hg18) and identified a total of 1,409,406 SNVs and short indels. However, only 163,502 (11.6%) variants had sufficient coverage to pass filtering (see Materials and Methods) and were included in the final list for the analysis.

In parallel to SNV discovery, we used the MACS algorithm (90) to map genome-wide ER binding sites using the same nine ER ChIP-seq data sets from above (but each data set analyzed independently). The results showed a wide range of variability in the number of

binding peaks from 15,677 to 79,978 sites. To build a consensus peak list, we overlapped the binding sites of all data sets and selected the genomic regions which were common in at least six data sets. This led to the detection of 22,143 ER binding sites with an average length of 385 bp. Using this panel of ER binding peaks, we next identified the SNVs which altered consensus EREs.

Motif assessment was performed by comparing ER binding probabilities in the presence and absence of SNVs. The variants that were associated with a statistically significant change (see Materials and Methods) were selected as putative regSNVs. Our pipeline nominated 5,839 motif altering regSNVs, among which 3,067 (53%) and 2,772 (47%) variants were computationally predicted to increase and decrease the binding affinity of their corresponding motifs, respectively (Table 10). To further refine the list, regSNVs were annotated with the closest adjacent genes and this list was compared to a list of estrogen-regulated genes. We focused on regSNVs capable of increasing ER binding and being within the proximity of an E2-regulated gene (<5 kb of distance) (Table 1). Interestingly, a number of highly ranked statistically significant putative regSNVs appeared close to genes previously shown to be oncogenic in breast cancer such as PVT1 (94), IGF1R (95) and GREB1 (96). Of these, rs62022087 located in *IGF1R* was identified by both JASPAR and TRANSFAC matrices increasing the confidence of the call. Moreover, Sanger sequencing showed that this regSNV is heterozygous in MCF7, which makes it an appropriate candidate for allele-specific binding assays. This prompted us to investigate regulatory function of rs62022087 by further *in vitro* studies.



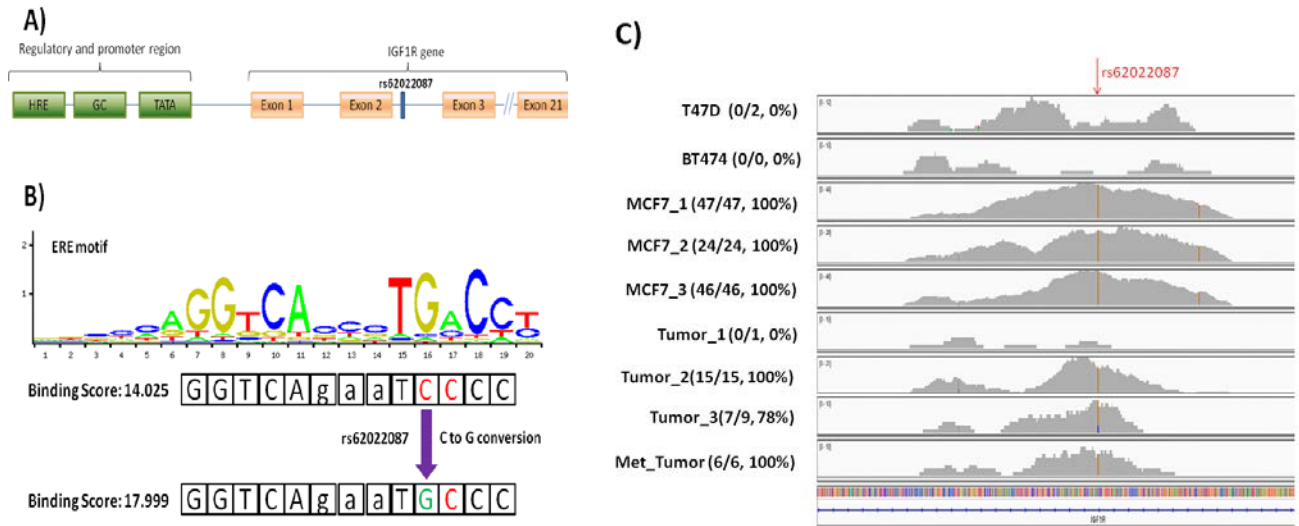
**Table 2. Top 10 regulatory SNVs increasing ER binding to MCF7 genome.**

Chr	Annotation	Gene	Distance From Gene	SNP ID	Adj_Pvalue_Ratio
chr8:128992864	ncRNA_intronic	PVT1	NA	NA	2.26E-06
chr10:94821513	intergenic	CYP26C1,CYP26A1	dist=3069;dist=1709	rs68040629	1.10E-05
chr15:97136484	intronic	IGF1R	NA	rs62022087	2.03E-05
chr15:97136484	intronic	IGF1R	NA	rs62022087	2.03E-05
chr10:121292409	upstream	RGS10	NA	rs10787978	3.39E-05
chr6:157157941	intronic	ARID1B	NA	rs12208040	3.63E-05
chr6:157157941	intronic	ARID1B	NA	rs12208040	3.63E-05
chr11:20014669	intronic	NAV2	NA	rs10741810	3.65E-05
chr17:54818764	intronic	YPEL2	NA	rs8073731	5.44E-05
chr2:10384622	intronic	HPCAL1	NA	rs2014889	5.62E-05
chr2:10384622	intronic	HPCAL1	NA	rs2014889	5.62E-05
chr4:3456949	intronic	DOK7	NA	rs916189	1.09E-04
chr2:11712184	intergenic	GREB1,NTSR2	dist=11821;dist=3571	rs6432223	1.13E-04

### 2.3.2 ER binding is associated with an intronic regSNV in *IGF1R*

Our motif assessment analysis showed that rs62022087 is one of the top three regSNVs putatively modulating ER binding to an ER-regulated gene. This SNV is located within an ERE, and the G of the SNV could potentially alter the ERE from a weak to a strong binding site. rs62022087, with a minor allele frequency (MAF) of 13.5%, is located centrally in the second intron of *IGF1R* (Figure 5) which is a region hosting several active histone marks such as H3K29ac and H3k4Me1, and a number of transcription factors including FOXA1, FOXA2 and E2F1, and finally DNase I hypersensitive sites (Figure 16). Direct genotyping of rs62022087 by Sanger sequencing of MCF7 genomic DNA indicated that the locus is heterozygous compared to T47D, ZR75 and BT474 cells. We examined whether ChIP-seq data showed an allelic preference towards the regSNV, as would be predicted from the increased ERE motif binding (4). Supporting this, cell lines (MCF7) and human breast tumors (Tumor\_2, Tumor\_3 and

Met\_Tumor) which harbor the regSNV showed increased ER ChIP-seq reads in this ER binding site (Figure 5). In addition, the allele frequency of rs62022087 is strongly biased towards the variant allele in the samples carrying the regSNV (MCF7: 100%, Tumor 2: 100%, Tumor 3: 78%, Met Tumor: 100%, derived from (4)), further supporting the idea that the regSNV results in increased ER binding. A similar phenomenon was observed in the ChIP-seq data sets of two other studies (Figure 17). Collectively, these data suggest that ER has higher affinity for the regSNV allele compared to the wild-type allele resulting in an increased transcriptional activity and expression of IGF1R.

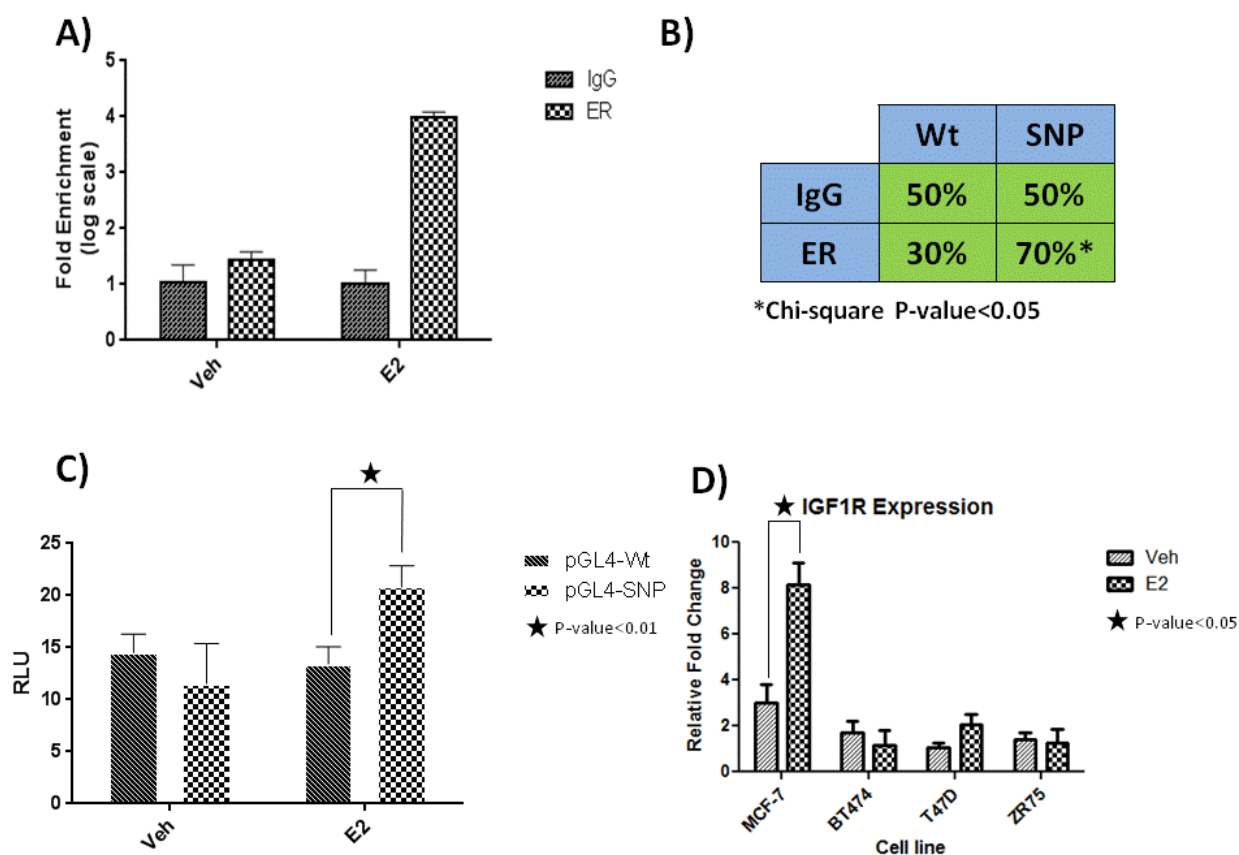


**Figure 5. The location of rs62022087 in genome and ER binding sites.**

A) Schematic view of SNV genomic position in IGF1R gene. B) The position of IGF1R SNP with regards to canonical ERE sequence. C) The distribution of ER ChIP-seq reads flanking rs62022087 SNP in different cell line models as well as patient tumors (ref. 1). The numbers in parentheses are the fraction and percentage of the reads containing mutant allele, respectively.

We next performed experiments to directly examine the role of the regSNV in altering ER-mediated induction of IGF1R expression. ER ChIP-qPCR in MCF7 cells showed that ER bound the genomic region containing regSNV in intron 2 of *IGF1R* with a 4-fold enrichment

following E2 treatment (Figure 6A). Allele-specific ChIP showed a significant enrichment of the regSNV allele (G allele) in the DNA bound to ER (Figure 6B). Cloning of the ER binding site (with or without the regSNV site) upstream of a heterologous promoter and luciferase indicated that the ER binding site containing the regSNV showed greater ER-induced luciferase expression upon estradiol treatment (Figure 6C). This indicates that the G allele is more potent in recruiting ER and subsequently transcriptional initiation leading to increased induction of IGF1R expression (Figure 6D). Consistent with this, we observed a significant increase in IGF1R transcript in MCF7 cells compared to the cell lines that lack the regSNV and are homozygous for the wild-type allele. Taken together, our *in vitro* experiments validate one of the top candidates found by our computational pipeline showing rs62022087 can change chromatin configuration in favor of ER binding and in the higher expression of IGF1R gene.



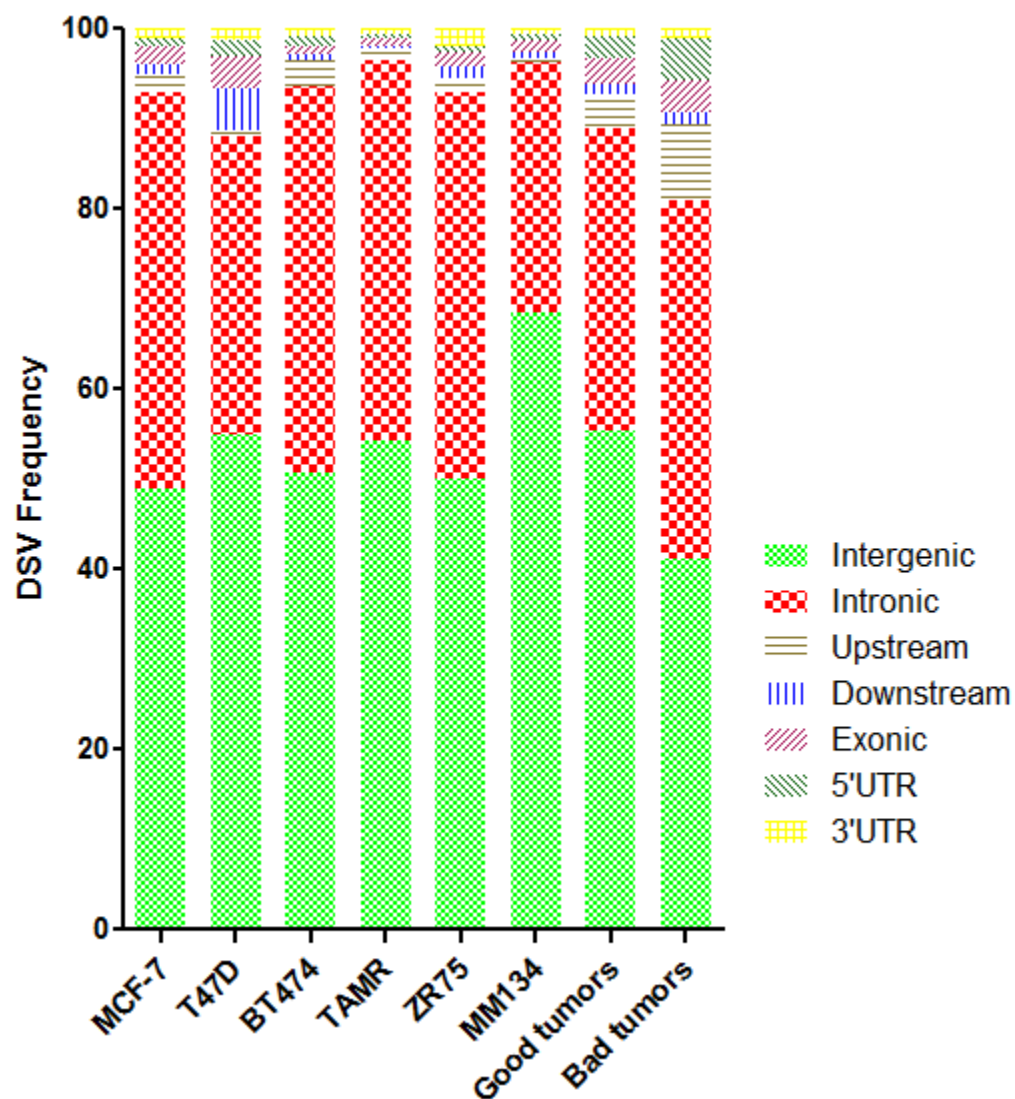
**Figure 6. IGF1R SNP can affect ER binding and result in higher gene expression.**

A) Confirmation of ER binding to IGF1R SNP by ChIP-qPCR in MCF7 cell line. The cells were estrogen deprived for 3 days and subsequently treated by Veh or E2 (1nM) for 45 minutes. ChIP was performed as describes in the methods section. ER binding is significantly enriched upon treatment by E2. B) Allele specific ChIP result shows a significant enrichment of SNP allele (70%) vs Wt allele (30%) in ER binding site. C) Luciferase transactivation assay using MCF7 cells transfected with constructs containing the ER binding site with Wt or SNP. The luciferase assay demonstrates that the binding site with variant allele has higher affinity to ER upon induction by Estradiol (1nM). D) IGF1R gene expression in different breast cancer cell lines treated by Veh or E2 (1nM). The significant induction of IGF1R expression in MCF7 Cell line may contribute to the presence of regulatory SNP compared to the other cell lines with WT allele.

### 2.3.3 Discovery of RegSNVs in available breast cancer ER ChIP-seq data

We applied our workflow to all available ER ChIP-seq data in breast cancer cell lines and tumors comprising a total of 43 data sets from 7 independent studies (Table 10-Table 18) (2-5,44-46,93,97). The variant calls were confined to those within ER binding sites. The closest genes to

the RegSNVs were annotated. The genomic position of the RegSNVs was also defined based on the coding and regulatory annotations. Figure 7 shows the distribution of RegSNVs in the analyzed models from available ER ChIP-seq data. As expected, the majority of regulatory variants are located in the intergenic areas whose functionality is not well-characterized. A great portion of the SNVs lies in intronic areas suggesting a major role of introns in regulation of the gene expression.



**Figure 7. The distribution of RegSNVs over the genome across a panel of breast cancer cell lines, good and bad prognosis tumors.**

The binding sites from different ER ChIP-seq data sets were extracted and annotated based on their location in the genome. The majority of the binding sites are located in the intergenic and intronic areas.

Using gene expression on 1,045 samples in TCGA, we found that regSNVs (n=11,605) are enriched in the proximity of genes differentially regulated between ER-positive (n=808) and ER-negative tumors (n=237) (chi-square test, pvalue<0.01). Further, to determine if the regSNVs have a functional role, we assessed the correlation of genotype (i.e. regSNV) with neighboring gene expression. We used the SNP array data in TCGA to find the samples carrying the SNVs and then compared the expression of target genes in SNV vs wild-type carriers in only ER+ samples. Interestingly, this led to the discovery of 17 regSNVs associated with the expression of their adjacent genes (qvalue<0.01, Table 3). We observed that the variant allele was enriched in the ER binding sites where there was at least coverage of 10x (Table 19). This indicates that higher affinity of ER to variant allele leads to higher expression of the target gene. The majority of these variants (13 out of 17) were located in the promoter of target genes further showing that they are likely functionally important regulatory variants (Table 3).

**Table 3. Top regulatory SNVs associated with the expression of their target genes**

RegSNV	Location	Target Gene	Number of Tumors with SNV genotype	log2 fold change (Mut/WT)	adj p-Value
rs36208869	Promoter	GSTM1	32	4.580808174	1.25E-08
rs1131017	Promoter	RPS26	318	-0.394006779	5.19E-07
rs7113753	Promoter	TRAPPC4	180	0.258426245	2.79E-05
rs1412825	Promoter	LRRIQ3	243	-0.217910733	3.64E-05
rs34282253	Promoter	XKR9	119	0.407692927	4.62E-05
rs10747783	Promoter	TSFM	205	-0.218721922	0.000157917
rs252923	Promoter	SETD9	197	0.407801595	0.000157917
rs41293275	Promoter	NSUN4	175	-0.215401706	0.000241865
rs3213745	Promoter	CEBPZ	241	-0.191620271	0.000444457
rs2732649	intergenic	LRRC37A	132	0.124517689	0.002214471
rs17361749	Promoter	NSUN4	168	-0.196072417	0.002736821
rs10489769	Promoter	NSUN4	172	-0.189003347	0.004515197
rs10956142	intergenic	ANXA13	38	-0.291199916	0.004515197
rs2939587	Promoter	TM2D3	260	0.207588052	0.005471564
rs1291363	Promoter	HTR7P1	315	0.591907017	0.006560413
rs4418583	Intron	LDLRAP1	248	0.237085892	0.006560413
rs3811254	Intron	OR4E2	3	0.035536509	0.009385423

The top candidate in our list is rs36208869 which is an SNV in the promoter of GSTM1 gene. Our algorithm predicted an increased binding of ER to the SNP allele and we observed an approximately 16 fold higher expression in tumors carrying the SNP (adjP=1.25E-08). GSTM1 encodes for a member of the glutathione S-transferase family which is responsible for detoxification of chemical compounds including carcinogens and products of oxidative stress (98). A large body of evidence has shown that loss of GSTM1 increases the susceptibility to several types of cancer including lung and bladder (99-101). Interestingly, we inquired Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) data and found that higher expression of GSTM1 in breast tumors is associated with better survival in a group of 1,505 patients (Figure 23, pvalue=8.2E-4). This result may be of clinical importance as rs36208869 may predict a better survival in patients carrying the variant allele.

#### **2.3.4 Identification of somatic RegSNVs in WGS data of breast tumors**

In order to discover novel recurrent somatic mutations in ER binding sites, we modified our workflow and applied it to a WGS study of 45 primary tumors paired with normal blood (88). The principles of our analysis follow what was described previously with minor modifications (Figure 18). We found 7,482 somatic SNVs occurring within ER binding sites. Among these variants, 13 were recurrent in more than one tumor (Table 4). We identified two intronic mutations in GPR126 and PLEKHS1 genes at a frequency of 8.6% (4/46). Interestingly, another mutation was observed 3 bp away in each locus at the lower frequency (2.1%, 1/46). We further validated these mutations in 98 breast tumors within TCGA samples with WGS data and, they appeared to be recurrent. However, the overall frequency of mutations is lower for both loci in TCGA (Figure 19). Based on previously published ChIP-seq data (4), the location of these

mutations overlap with ER binding further suggesting a regulatory role (Figure 19). In addition, somatic SNVs lie in regions shown to be actively regulating transcription by ENCODE data (Figure 20). We also found that non-coding mutations in GPR126 and PLEKHS1 are located within palindromic sites which are known to form sites for transcription factors binding (Figure 21). This observation was intriguing although we failed to find a correlation between mutations and gene expression of GPR126 and PLEKHS1 genes (Figure 22).

**Table 4. Frequent SNVs in putative ER binding sites within WGS of 45 tumors**

Chr	Start	End	Frequency (out of 46 tumors)	Closest gene
6	1.43E+08	1.43E+08	4	GPR126
10	1.16E+08	1.16E+08	4	PLEKHS1 (c10orf81)
8	99487670	99487671	3	KCNS2
3	1.04E+08	1.04E+08	3	ZPLD1
3	75824685	75824686	2	LINC00960
21	9757052	9757053	2	TPTE
21	10062034	10062035	2	BAGE2
16	80440896	80440897	2	PLCG2
16	10627835	10627836	2	TEKT5
15	36396249	36396250	2	SPRED1
14	1.01E+08	1.01E+08	2	LINC00524
1	2.06E+08	2.06E+08	2	CR1L
1	1.99E+08	1.99E+08	2	CACNA1S

## 2.4 DISCUSSION

Global genetic variation in TF binding sites can lead to widespread changes in gene expression among different individuals (102-104). Analyzing complete genomes of different cancer types has elucidated recurrent mutations in the genomic regions potentially regulated by TFs (81,82,105). However, deciphering how genome-wide DNA variants affect TF binding remains



understudied. We present a computational pipeline, which analyzes ChIP-seq reads to identify regSNVs in TF binding sites. We used this pipeline, in combination with experimental studies, to confirm the impact of regSNVs on the corresponding DNA motifs. While other studies have identified regSNVs in ER binding sites using a biased approach involving genotyping information from resources such as dbSNP and GWAS, our approach differs by identifying SNVs directly from ChIP-seq data thus increasing the likelihood of identifying regSNVs in TF binding sites..

The MCF7 cell line is one of the most studied models for understanding ER biology, and results from this cell line have had a fundamental impact upon breast cancer research and patient outcome (67). Using available ER ChIP-seq data in MCF7, we investigated the genetic variation in ER binding sites with this model. The number of binding sites varies significantly between the MCF7 data sets ranging from 15,677 to 79,978 sites. This high degree of variation may be due to slight differences in technical details, such as culturing conditions or cell line passage numbers, utilized for the ChIP experiments. We used an overlap of ER binding sites for this study. Our analysis revealed a functional regSNV (rs62022087) in intron 2 of the *IGF1R* gene which was predicted to increase ER binding. We show that the rs62022087 SNP results in increased ER recruitment to intron 2 and increased E2-mediated expression of IGF1R gene in MCF7 cells compared to cell lines carrying the wild type allele. IGF1R overexpression has been implicated to play an important role in the development of breast cancer (106-108) and the crosstalk between IGF1R and estrogen signaling has been well established in malignant breast tissue (38,109,110). This prompted us to obtain more information on this SNP from GWAS studies and correlate it with clinical outcome in breast cancer patients. However, neither rs62022087 nor any of the SNPs in LD with our candidate SNP are genotyped by Affymetrix chips, which are

commonly, used in GWAS studies. Further sequencing studies in large cohorts are warranted to characterize the potential role of this regulatory SNP in development and progression of breast cancer.

Our pipeline is able to detect not only germline variants but also rare somatic mutations, which may alter the affinity of TF to DNA. However, the general low coverage of ChIP-seq data makes it challenging to perform accurate variant calling. Therefore, in this study we pooled the reads from multiple data sets on the same cell line to improve the confidence of calls. With the decreased costs of sequencing, we expect that increased coverage in ChIP-seq studies will alleviate this problem in the near future.

Applying our pipeline to all available ER ChIP-seq data characterized thousands of RegSNV candidates in multiple breast cancer models, which may potentially change the binding of ER. About 96% of these variants are annotated in dbSNP and 1000genome databases and are thus likely to be germline alteration, but we didn't have access to normal matched samples to confirm this. This high rate of germline SNPs may reflect our inability to detect low allele frequency somatic mutations due to the low read coverage of ChIP-seq data. The majority of regSNVs reside in intronic regions of the genome, similar to the regSNV we have characterized in intron 2 of the *IGF-IR* gene. Several studies have identified regulatory SNPs in genes associated with breast cancer susceptibility and treatment (49-51,111). By integrating multi-omics large data sets, we found 17 regSNVs associated with the expression of target genes. The Top candidate was a SNP in the promoter of *GSTM1* whose expression is associated with survival in breast cancer patients. ChIP-seq reads provided further evidence showing the variant allele is enriched in the ER binding sites although we were not able to infer the true reference genotype due to not having access to normal tissue information in analyzed samples (Table 19).

We finally sought to identify recurrent non-coding mutations in ER binding sites by integrating WGS and ChIP-seq data. We discovered two mutation hotspot regions in the introns of GPR126 and PLEKHS1 genes. Several studies have previously reported the recurrence of PLEKHS1 mutations not only in breast but also in other cancers (81,82,105). The biology of these two genes is not known in cancer. GPR126 is a G-protein coupled receptor that is involved in neural development and myelination in mammals (112,113). PLEKHS1 is a pleckstrin homology domain containing protein which has been shown to be regulated by E2 in MCF7 cell line (114). We also did not find any meaningful connection between the presence of mutations and expression of neighboring genes in TCGA data. Therefore, we suspect these hotspots may interact with distal target genes. Future experiments such as chromosome conformation capture combined with high-throughput sequencing (4C-seq) are required to screen for potential physical interaction of these regulatory elements and distal parts of the genome. This will help us identify non-coding mutations affecting the regulation of critical genes in cancer.

### **3.0 SENSITIVE DETECTION OF MONO- AND POLYCLONAL *ESR1* MUTATIONS IN PRIMARY TUMORS, METASTATIC LESIONS AND CELL FREE DNA OF BREAST CANCER PATIENTS\***

\*This chapter has been published in Clinical Cancer Research (67) and permission has been obtained from the journal regarding the copy-right. CCR correspondence: Karola Rac (karola.rac@aacr.org)

Peilu Wang<sup>1,2,\*\*</sup>, Amir Bahreini<sup>2,3,\*\*</sup>, Rekha Gyanchandani<sup>2,4,\*\*</sup>, Peter C. Lucas<sup>5</sup>, Ryan J. Hartmaier<sup>2,4</sup>, Rebecca J. Watters<sup>2,4</sup>, Amruth R. Jonnalagadda<sup>2</sup>, Humberto E. Trejo Bittar<sup>5</sup>, Aaron Berg<sup>5</sup>, Ronald L. Hamilton<sup>5</sup>, Brenda F. Kurland<sup>6</sup>, Kurt R. Weiss<sup>7</sup>, Aju Mathew<sup>8</sup>, Jose Pablo Leone<sup>8</sup>, Nancy E Davidson<sup>2,8</sup>, Marina N. Nikiforova<sup>5</sup>, Adam M. Brufsky<sup>2,8</sup>, Tadeu F. Ambros<sup>8</sup>, Andrew M. Stern<sup>9</sup>, Shannon L. Puhalla<sup>2,8</sup>, Adrian V. Lee<sup>2, 3, 4\*\*</sup>, Steffi Oesterreich<sup>2, 4\*\*</sup>

<sup>1</sup> School of Medicine, Tsinghua University, Beijing, PRC, 100084

<sup>2</sup> Womens Cancer Research Center, Magee-Women Research Institute

<sup>3</sup> Department of Human Genetics, University of Pittsburgh

<sup>4</sup> Department of Pharmacology and Chemical Biology, University of Pittsburgh

<sup>5</sup> Department of Pathology, University of Pittsburgh

<sup>6</sup> Department of Biostatistics, University of Pittsburgh

<sup>7</sup> Department of Orthopedic Surgery, University of Pittsburgh Medical Center (UPMC)  
Pittsburgh

<sup>8</sup> Department of Medicine, Division of Hematology/Oncology, University of Pittsburgh Cancer  
Institute and UPMC CancerCenter

<sup>9</sup> Drug Discovery Institute, University of Pittsburgh  
Pittsburgh, PA, USA, 15213

\*\* PW, AB, and RG contributed equally to this work, and AVL and SO share senior authorship

### **Author's Contribution**

Conception and design: A. Bahreini, R.L. Hamilton, K. Weiss, A. Mathew, N.E. Davidson, A.M.  
Brufsky, T.F. Ambros, S. Puhalla, A.V. Lee, S. Oesterreich

Development of methodology: P. Wang, A. Bahreini, R. Gyanchandani, R.J. Hartmaier, A.R.  
Jonnalagadda, K. Weiss, A. Mathew, J.P. Leone, M.N. Nikiforova, A.M. Brufsky, S. Puhalla,  
A.V. Lee, S. Oesterreich

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):  
P. Wang, R. Gyanchandani, R.J. Watters, A.R. Jonnalagadda, H.E.T. Bittar, A. Berg, R.L.  
Hamilton, K. Weiss, A. Mathew, J.P. Leone, N.E. Davidson, M.N. Nikiforova, A.M. Brufsky,  
T.F. Ambros, S. Puhalla, A.V. Lee

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational  
analysis): P. Wang, A. Bahreini, R. Gyanchandani, P. Lucas, R.J. Hartmaier, A.R. Jonnalagadda,

B.F. Kurland, K. Weiss, A. Mathew, J.P. Leone, M.N. Nikiforova, A.M. Brufsky, T.F. Ambros, A.M. Stern, S. Puhalla, A.V. Lee, S. Oesterreich

Writing, review, and/or revision of the manuscript: P. Wang, A. Bahreini, R. Gyanchandani, P. Lucas, R.J. Hartmaier, R.J. Watters, A.R. Jonnalagadda, H.E.T. Bittar, A. Berg, R.L. Hamilton, B.F. Kurland, K. Weiss, A. Mathew, J.P. Leone, N.E. Davidson, A.M. Brufsky, T.F. Ambros, A.M. Stern, S. Puhalla, A.V. Lee, S. Oesterreich

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): A. Bahreini, R. Gyanchandani, R.J. Watters, A.R. Jonnalagadda, A. Mathew, A.V. Lee, S. Oesterreich

Study supervision: K. Weiss, A.M. Stern, S. Oesterreich

### 3.1 INTRODUCTION

ER is expressed in the majority of breast cancers and is a major regulator of breast cancer development and progression (115). Endocrine therapy is one of the most efficacious and least toxic treatments in ER-positive (+) breast cancers. Current strategies target ER action either by ligand deprivation (AIs or ovarian function suppression) or ER blockade through selective SERMs and SERDs. All these therapies may improve survival in early stage breast cancer (22,23). However, *de novo* or acquired resistance is a major clinical problem, especially in metastatic breast cancer. Multiple molecular mechanisms of resistance include down-regulation of ER expression, dysregulation of ER co-regulators, post-translational modifications of ER, and crosstalk with growth factor signaling pathways (6,34-39,116).

The concept that somatic base-pair missense mutations in *ESR1* may confer hormone independence has been speculated for many years. However, studies of primary breast cancer have reported few or no *ESR1* mutations (53-57). For example, *ESR1* base-pair missense mutations are present at 0.2% (1/482) in breast cancers in TCGA (111), and 0.3% (5/1430) in the Catalog of Somatic Mutations in Cancer. However, recent studies have documented *ESR1* as being highly mutated in metastatic breast cancer. Li *et al* first reported *ESR1* ligand binding domain mutations in two patient-derived xenografts from hormone-resistant advanced disease (59). Subsequently, high rates of *ESR1* mutation (15-50%) in metastatic breast cancer have been reported (60-63). Furthermore, recent studies have implicated that the emergence of *ESR1* fusions can also be a mechanism of endocrine therapy resistance (59,117). Preliminary functional studies indicate that some somatic mutations in *ESR1* results in ER ligand-independent activity that is partially resistant to current endocrine therapies, suggesting that these mutations may undergo selection under the pressure of endocrine therapy (59-63).

One goal of precision cancer medicine is to make clinical decisions based upon genomic data, which can identify a target for therapy, and/or predict therapeutic resistance. It is hypothesized that *ESR1* gene mutations may be a predictive biomarker of resistance to endocrine therapy. As longitudinal biopsy and genetic analysis of metastatic disease is often not feasible, the concept of measuring mutations in tumor DNA circulating in plasma, termed cfDNA, has recently gained much attention. The feasibility of using cfDNA to noninvasively identify molecular alterations within metastatic tumors has been shown in several studies (68,118,119) and preliminary data suggest that cfDNA can be used to monitor breast cancer burden and treatment response (120). A recent proof-of-principle study detected an *ESR1* mutation (E380Q) in cfDNA from a single patient with advanced hormone refractory breast cancer (68,69). However, the detection of rare mutations has been challenged by several limiting factors, including low cfDNA yields and low tumor cellularity in metastatic lesions. ddPCR is a highly sensitive and robust technology for detection of rare mutations compared to the available sequencing techniques (119,121,122). Here, we report the use of ddPCR to study the incidence of *ESR1* mutation in primary breast cancer, metastatic biopsies with a focus on bone and brain metastases since they have been understudied due to difficulties in accessing such tissue, and finally cfDNA from breast cancer patients with recurrent disease.



## **3.2 METHOD**

### **3.2.1 Sample acquisition**

Samples used in this study were obtained from the University of Pittsburgh Health Sciences Tissue Bank (HSTB) (primary breast cancer, brain metastases), or were prospectively collected (bone metastases, blood). There were no special criteria for selection of samples for the study other than those described here. Frozen primary ER-positive breast cancers (n=43) (>60% tumor cellularity) from patients subsequently treated with endocrine therapy were obtained from HSTB. Metastatic tumor biopsies from brain (n=38) and bone (n=12) were collected through HSTB over the last three years. For collection of cfDNA (n=29), blood was drawn (1-4 x 10ml Streck tubes) between 01/14 and 08/14 from patients with advanced disease seen within the UPMC health system. There were a total of 122 samples, from 121 patients, since one patient (CF28) donated both cfDNA and a bone metastases sample. In addition, we had access to skin (CF4), liver (CF16), ovarian (CF23), and soft tissue neck metastases (CF14) from patients who donated blood for cfDNA isolation, thus totaling n=126 analyzed samples. ER status was detected by immunohistochemistry, using ASCO-CAP 2010 guidelines for tumors diagnosed in or after 2010 (123). All patients signed informed consent, and the studies were approved by the University of Pittsburgh IRB.

### **3.2.2 DNA isolation, preparation, and quantification**

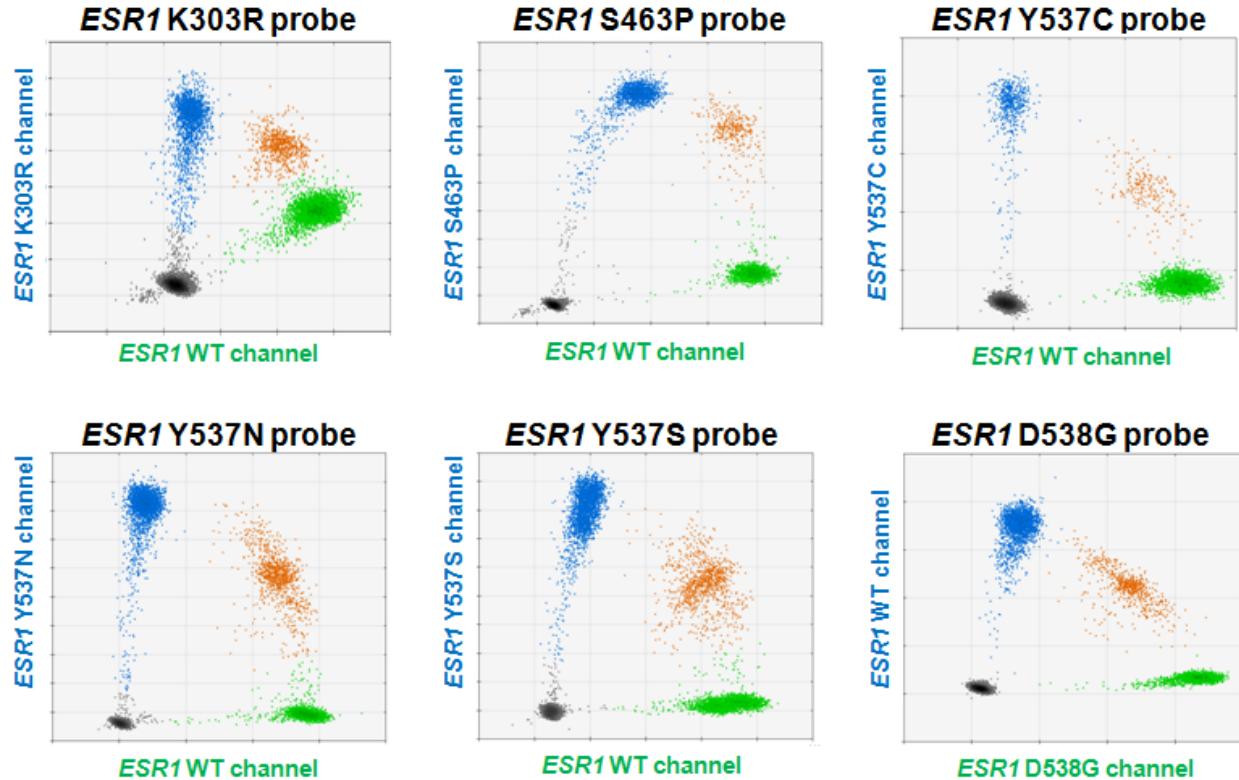
30-50 mg of frozen primary tumor tissue and 50-150 mg frozen bone metastases were crushed under liquid nitrogen, and DNA was isolated using Qiagen DNeasy Blood & Tissue Kit. Brain

metastases were obtained as FFPE sections and Qiagen Allprep DNA/RNA FFPE Kit was used to isolate DNA from four to six 10 µm slides. cfDNA was isolated as previously described (121). Briefly, plasma was separated by double centrifugation within 7 days of blood collection, and DNA was isolated from 1-4 ml plasma using QIAamp Circulating Nucleic Acid kit. Targeted high-fidelity pre-amplification (15 cycles) was performed on cfDNA and DNA isolated from FFPE brain metastases using primers listed in Table 20. Pre-amplification products were purified using QiaQuick PCR purification kit and diluted before ddPCR at 1:100 and 1:20 for brain metastases and cfDNA, respectively. The pre-amplification does not affect linearity of detection of the mutant allele, as we have shown for *ESR1* and PIK3CA mutations (Figure 24 and Figure 25). All DNA samples were quantified by Qubit dsDNA HS/BR assay kits (Life Technologies).

### **3.2.3 Mutation detection by ddPCR**

Primers and probes were designed and ordered through Life Technologies for S463P, Y537C/N/S, K303R and Integrated DNA Technologies for D538G *ESR1* mutations (Table 21). Bio-Rad QX100 Droplet Digital PCR system was used. Briefly, 1 µl template from diluted pre-amplified products or 50-60 ng of non-amplified DNA was mixed with ddPCR supermix for probes (no dUTPs) (Bio-Rad laboratories, Inc.) and primer/probe set. Droplets were generated using 20 µl of the reaction mixture and 70 µl of droplet generation oil. Positive and negative controls were included in each run to exclude potential contamination artifacts, and to control for proper gating of alleles. All mutation-positive samples were run in at least 3 replicates, assaying at least 10,000 genome equivalents. For positive controls, we utilized oligonucleotides containing the mutation (463P, Y537C/N, K303R), DNA from a cell line with a D538G knock-in

mutation (unpublished data), or DNA from a liver biopsy with an *ESR1* mutation at Y537S confirmed by Sanger sequencing (Figure 8). Specificity of the probes was demonstrated for Y537C/N/S and D538G mutations (Figure 26). No detectable cross-reactivity of mutant probes and WT probes was observed for D538G or Y537C mutation (Figure 27). We did find that an increase in the presence of Y537S caused a slight downshift in the fluorescent signal for D538G (Figure 28) causing a double population, however this didn't affect the calculated D538G allele frequency. The reason for the decrease in D538G fluorescence is unclear. Mutations with high allele frequencies were confirmed with Sanger sequencing using primers listed in Table 20.



**Figure 8. Positive controls for mutation probes utilized in ddPCR technology.**

*ESR1* K303R, S463P, Y537C, Y537N oligos, or *ESR1* D538G and Y537S gDNA were mixed with *ESR1* WT gDNA to serve as positive controls for the assay. Scatter plots of ddPCR results showing fluorescent detection of individual droplets. Blue and green dots represent droplets with *ESR1* genotypes indicated on Y-axis and X-axis, respectively. Orange dots represent droplets containing both WT and mutant *ESR1* DNA. Black dots represent droplets that did not contain DNA.

### 3.2.4 Quantitative analysis

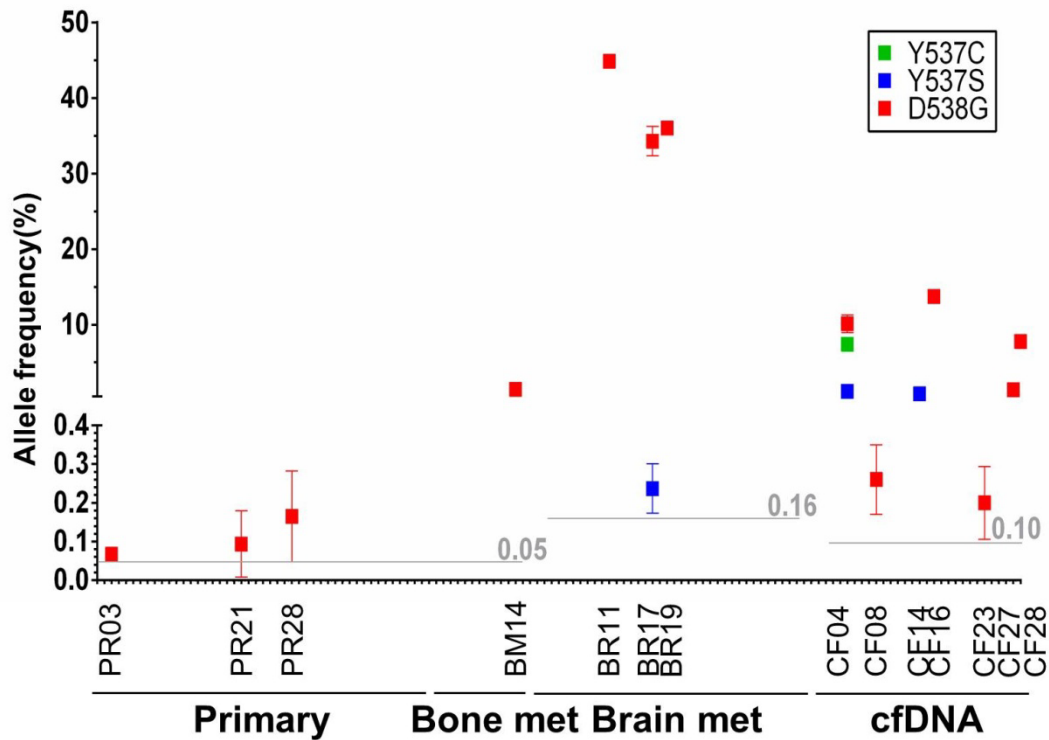
Data were analyzed using QuantaSoft software (Biorad), calculating a fractional abundance (“mutant allele frequency”). The background noise, which was higher in pre-amplified DNA from cfDNA and FFPE brain metastases compared to DNA from frozen tissues (primary tumors and bone metastases), was defined as the average of allele frequency plus half (for cfDNA) or full (for FFPE DNA) 95% confidence intervals (CIs) of negative controls (*ESR1* wildtype DNA) across all ddPCR assays. The noise was subtracted from the allele frequencies. The background-noise-adjusted lower limits of detection (LLoD) of the assay were 0.05% for frozen tissues, 0.10% for cfDNA, and 0.16% for FFPE tissues (Figure 29). Samples were called “positive” for the *ESR1* mutation if a) the allele frequencies were >0 after subtraction of background noise, AND b) >2 mutant droplets were repeatedly detected, AND c) allele frequency was > noise adjusted LLoD for at least 3 independent assays.

## 3.3 RESULTS

### 3.3.1 *ESR1* mutations in primary tumors

We screened 43 primary ER-positive tumors to detect *ESR1* mutations (S463P, Y537C, Y537N, Y537S, and D538G) recently described in recurrent endocrine-resistant breast cancer. We also included the analysis of the K303R mutation, which has been previously described to be present in primary and metastatic disease, while it wasn’t detected in other studies (74,124-127). Three primary tumors (PR3, PR21, PR28) were positive for D538G, with very low mutant allele

frequencies between 0.07 to 0.2% (Figure 9, and Table 5). Another sample (PR44) was positive in multiple repetitive assays, but the mutant allele frequency (0.012%) was below our LLoD. No other mutations were detected in any of the remaining primary tumors. We thus detected *ESR1* mutations in 7.0% (3/43, 95% Wilson binomial confidence interval (CI) 2%-19%) of primary ER+ breast cancers.



**Figure 9. *ESR1* mutation allele frequency of *ESR1* mutation-positive samples.**

Average mutant allele frequency  $\pm$  SEM were calculated using data from at least 3 replicates (after subtraction of respective background noise). Grey lines indicate the adjusted LLoD of respective tissue. PR, BM, BR, and CF represented primary tumors, bone metastases, brain metastases, and cfDNA respectively. Each mark on x-axis represents a sample, and names are indicated for *ESR1* mutation-positive samples.

**Table 5. The rates of *ESR1* mutations in primary tumors, cfDNA, brain and bone metastases from breast cancer patients.**

Samples	N	ER+ Primary	<i>ESR1</i> mutations						Pts with <i>ESR1</i> mutation	Rates of <i>ESR1</i> mutation
			K303R	S463P	Y537C	Y537N	Y537S	D538G		
Primary tumor	43	43	0	0	0	0	0	3	3	7.0% (3/43)
Bone metastases	12	11 <sup>‡</sup>	0	0	0	0	0	1	1	8.3% (1/12) (9.1% in ER+)**
Brain metastases	38	24 <sup>‡</sup>	0	0	0	0	1	3	3*	7.9% (3/38) (12.5% in ER+)**
cfDNA	29	29 <sup>‡</sup>	0	0	1	0	2	6	7*	24.1% (7/29)

<sup>‡</sup> Number with known ER+ primary tumor at time of diagnosis (ER status of primary tumor unknown for remaining samples)

\* One patient with a brain metastasis, and one with cfDNA analysis had multiple distinct *ESR1* mutations within a single sample (polyclonal *ESR1* mutations).

\*\* Frequency of mutations in metastases from a known ER+ primary tumor

### 3.3.2 *ESR1* mutations in bone metastases

Since decalcification of bone metastases can impact downstream analyses, we restricted our analysis of bone metastases to fresh frozen tissue. We obtained 12 frozen bone metastases, 11 of which were from primary tumors known to be ER+. One sample (BM14) was positive for the D538G mutation, with an allele frequency of 1.4% (Figure 9), for an overall *ESR1* mutation rate in bone metastases of 8.3% (1/12, 95% CI 0.4%-35%). Of note, the pathologist's estimate of tumor cellularity in this sample was about 1-5% (Table 22), suggesting that the allele frequency of this mutation within tumor cells in this sample is likely much higher. To confirm our ability to detect mutations across all samples, we performed an additional control by assaying for a frequent *PIK3CA* mutation (H1047R). Three samples (BM01, BM08, and BM11) (Table 22) tested positive for *PIK3CA* mutation at high allele frequencies (27.0%, 29.7%, and 37.8%), supporting suitability of our metastatic samples for mutation detection by ddPCR (Figure 30).

### 3.3.3 *ESR1* mutations in brain metastases

We analyzed 38 brain metastases, 24 of which originated from ER+ primary tumors, and 14 from which the ER status of the primary tumors were unknown. All brain metastases with the exception of BR55 (30-40%), BR56 (40-60%), BR60 (40%) and BR68 (30-50%) had at least 60% tumor cellularity. Three brain metastases (BR11, BR17, BR19) contained D538G mutations at high allele frequencies (34.3-44.9%) (Figure 9 for an overall mutation rate of 7.9% (3/38, 95% CI 3%-21%): all were recovered from patients with ER+ primary breast cancer giving a 12.5% frequency in disease with known ER-positivity (3/24, 95% CI 4%-31%). The presence of the D538G mutation was confirmed by Sanger sequencing in the three brain metastases (Figure 31). Interestingly, sample BR17 had an additional Y537S mutation at a lower allele frequency (0.24%). Furthermore, using a dual-mutation specific probe, the mutations were found to be on separate alleles, indicative of polyclonal *ESR1* mutations within a single metastatic tumor (Figure 32).

### 3.3.4 *ESR1* mutations in cfDNA

We next interrogated *ESR1* mutations in cfDNA collected from 29 patients with metastatic breast cancer, all arising from ER+ primary disease. *ESR1* D538G (n=6), Y537S (n=2), and Y537C (n=1) mutations were detected in a total of 7 patients, with one patient (CF4) having polyclonal *ESR1*-mutations consisting of Y537C, Y537S and D538G with allele frequencies of 2.7%, 1.2% and 5.1% respectively (Table 5). cfDNA allele frequency was overall higher compared to primary tumors (Figure 9). The *ESR1* mutation rate in cfDNA was 24.1% (7/29, 95% CI 12%-42%).

Table 6 summarizes clinical characteristics and endocrine treatment history of patients with an *ESR1* mutation identified in cfDNA. Typical of patients with ER+ metastatic breast cancer, most had an extensive history of endocrine therapy as measured by both number of agents and months of exposure. There were not sufficient number of samples to formally analyze a predicted association between *ESR1* mutations and shorter survival.

**Table 6. Clinical characteristics and endocrine treatment history in patients with confirmed *ESR1* mutant cfDNA, brain or bone metastases.**

Clinical Characteristics							Endocrine Therapy before to Mutation Analysis			Endocrine Therapy after Mutation Analysis		
ID	Specimen	Detected <i>ESR1</i> mutations	Stage at Dx	ER Status	ADJ Hormonal therapy	ADJ Hormonal Therapy Duration (months)	Number of therapies	Cumulative Exposure (months)	Endocrine therapy	Number of therapies	Cumulative Exposure (months)	Endocrine Therapy
CF4 <sup>†</sup>	cfDNA	Y537C/S, D538G	IIB	+	SERM	5	3	23	AI, SERM	No	No	No
CF8 <sup>†</sup>	cfDNA	D538G	IIB	+	AI	13	1	47	SERD	2	5	AI, SERM
CF14 <sup>†</sup>	cfDNA/Soft tissue	Y537S	IV	+	No	0	4	25	AI, SERM, SERD	1	2	AI, SERM, SERD
CF16	cfDNA/Liver	D538G	0	+	No	0	4	35	AI, SERM, SERD	1	4	AI
CF23	cfDNA	D538G	IV	+	No	0	3	42	AI, SERD	1	7	SERM
CF27 <sup>†</sup>	cfDNA	D538G	IV	+	No	0	7	37	AI, SERM, SERD	No	No	No
BR11	Brain	D538G	0	+	No	0	4	47	AI, SERM, SERD	1	7	SERM
BM14/CF28	cfDNA/Bone	D538G	IIA	+	AI	72	2	9	AI, SERD	1	4	AI
BR19	Brain	D538G	NK	NK	NK	NK	NK	NK	NK	NK	NK	NK
BR17	Brain	Y537S, D538G	NK	NK	NK	NK	NK	NK	NK	NK	NK	NK

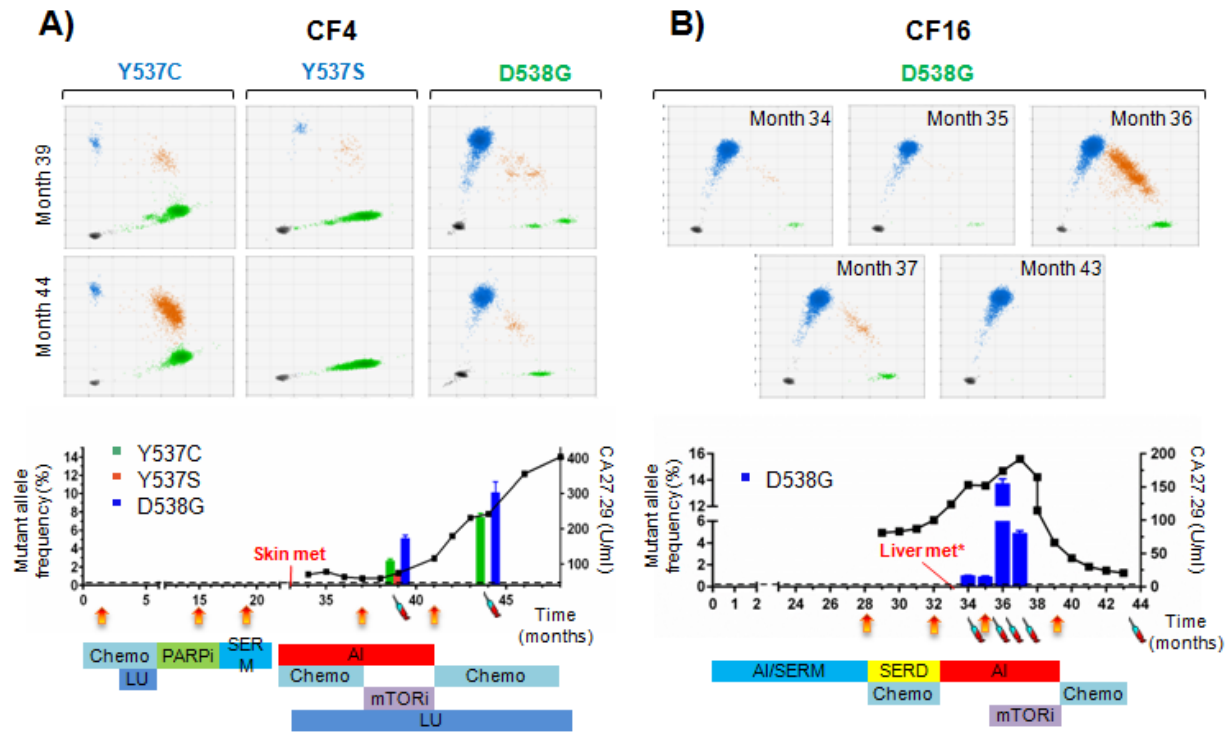
<sup>†</sup> Patient is deceased.

AI, aromatase inhibitor; SERM, selective estrogen receptor modulator; SERD, selective estrogen receptor down-regulator; ADJ, adjuvant; NK, not known; Dx, diagnosis



### 3.3.5 Analysis of *ESR1* mutations in serial blood samples, and matched metastatic tumors

Serial blood draws from four patients were available for longitudinal examination of *ESR1* mutation status. Patient CF4 (Figure 10A) was originally diagnosed with ER+ lymph-node positive disease, underwent mastectomy, and was then treated with SERMs. Over the next year, she developed metastases to brain, liver, bone and skin. A metastatic skin lesion biopsy was negative for *ESR1* mutation. A blood draw 6 months later showed three *ESR1* mutations with different allele frequencies (Y537C – 2.7%, Y537S – 1.2%, D538G – 5.1%). The patient received an aromatase inhibitor, everolimus, and chemotherapy for six months. A subsequent blood draw (6 months after the first one) revealed an enrichment of Y537C and D538G mutations, but a loss of the Y537S mutant clone (Y537C – 7.4%, Y537S < LLoD, D538G – 10.1%). The increase in the allele frequencies of D538G and Y537C co-occurred with an increase in the tumor marker CA 27-29.



**Figure 10. Clinical timelines and allele frequency of *ESR1* mutations in serial blood draws and matched metastatic lesions.**

A) Sample CF4 and B) Sample CF16. Top: 2D blots showing *ESR1* mutant allele frequency as measured by ddPCR. Bottom: The timeline starts with diagnosis of metastatic disease and shows treatments received, disease progression (indicated with orange/red vertical arrows), tumor marker assessments (CA 27-29 antigen line graph), blood draws (indicated with syringe), and *ESR1* mutant allele frequency (bar graphs). Treatment abbreviations: Chemo (chemotherapy), PARPi (PARP inhibitor), LU (Leuprolide), SERM (Selective Estrogen Receptor Modulator), SERD (Selective Estrogen Receptor Degradar), AI (Aromatase Inhibitor) and mTORi (mTOR inhibitor). \*The matched metastatic lesion was positive for *ESR1* mutation.

For patient CF16, DNA from five serial blood draws and from a biopsy of a liver metastasis was analyzed (Figure 9B). The patient originally developed ER+ chest wall metastases twelve years after excision of DCIS. She received serial endocrine therapy including tamoxifen, fulvestrant, and multiple AIs, followed by mTOR inhibitor and chemotherapy, but metastases progressed to other sites, including liver and bone. The *ESR1* D538G mutation was detected in both the liver metastasis (23.0%) and the first blood draw (1.0%). The allele frequency was similar in the 2<sup>nd</sup> blood draw (0.9%), peaked around the time of the 3<sup>rd</sup> draw (13.7%), decreased in the 4<sup>th</sup> blood draw (4.9%), and was below LLoD in the 5<sup>th</sup> draw taken after ~ 6 months of chemotherapy (0.2% before noise subtraction, which did not pass the cut-off for

“positive” mutation calling). The decreased frequency of the mutant allele corresponded to lower CA 27-29 levels after chemotherapy.

Two additional patients (CF23, CF28) had two blood draws each (Figure 33). Patient CF23 presented with Stage IV disease, with multiple bone lesions, and an ovarian metastasis that was negative for *ESR1* mutation. Blood was drawn at two time points throughout disease progression, as indicated in Figure 33A, which was approximately 1 month after surgical removal of the ovarian metastasis. D538G mutation was detected at low allele frequency (0.2%) in the 1<sup>st</sup> draw, and was below LLoD in the 2<sup>nd</sup> draw. Patient CF28 developed lung, bone, and brain metastases 3 years after completion of 5 years of AI treatment for an ER+ breast tumor (Figure 33B). She was treated with AI, and fulvestrant, and cfDNA from 1<sup>st</sup> blood draw was negative for *ESR1* mutations. The disease progressed, and a subsequent bone biopsy revealed an *ESR1* D538G mutation (1.4% allele frequency) (BM14, described above), and cfDNA showed the D538G mutation at 7.8% allele frequency. The increase in allele frequency of D538G co-occurred with an increase in CA 27-29 tumor marker. Finally, in one additional patient (CF14) with a single blood draw, the Y537S mutation was detected in both a posterior neck soft tissue nodule (40.5% allele frequency), and in cfDNA, although at lower frequency (0.8%).

Thus, in summary, mutations were either detected in both metastatic biopsy and cfDNA (n=3) or in cfDNA only (n=2), suggesting cfDNA as a source for disease phenotyping (e.g. detecting types of mutations), and potentially monitoring burden. This is supported by the observation that changes in *ESR1* mutation frequency correlated with changes in CA27-29 levels.

## 3.4 DISCUSSION

### 3.4.1 *ESR1* mutations are present at very low allele frequency in primary ER-positive breast cancer

Previous studies have shown low or undetectable rates of *ESR1* mutation in primary breast cancer using Sanger sequencing or massively parallel sequencing (MPS). This is the first study to examine *ESR1* mutations (S463P, Y537C, Y537N, Y537S, and D538G, K303R) in primary breast cancer using ddPCR. We found that 7.0% (3/43) of primary breast cancers have an *ESR1* D538G mutation, but the allele frequency is very low (0.07 to 0.2%). A recent NGS study of primary tumors from BOLERO trial identified *ESR1* mutation in 6/183 tumors (3.3%) (63). TCGA did not detect *ESR1* D538G mutation (or K303R, S463P, Y537C, Y537N, Y537S) in 482 primary breast cancers, and COSMIC contains only one *ESR1* D538G mutation from 1430 primary breast cancers. The very low allele frequency suggests that in some primary tumors *ESR1* mutations pre-exist as rare clones, which are then selected for during metastatic progression. This is consistent with a previous study from a single patient, which used deep-targeted MPS and identified an *ESR1* mutation (E380Q) at 2% allele frequency in primary disease and 68% in synchronous liver metastasis (68,69). Detection of rare *ESR1* mutations in primary tumors (0-7%) may be clinically relevant for predicting resistance to hormone therapy; however, additional studies using sensitive detection technologies are necessary to develop this area of investigation.

### 3.4.2 *ESR1* is mutated in both brain and bone metastases

Our analysis of 38 brain and 12 bone metastases showed *ESR1* mutations with higher allele frequency compared to primary tumors. To our knowledge, this is the largest study of *ESR1* mutations in these specific metastatic sites and the only one to use ddPCR. The most frequently identified *ESR1* mutation was D538G, which is consistent with five prior studies that detected a total of fourteen D538G mutations, eleven Y537S mutations, four Y537N mutations, three Y537C mutations, two S463P mutations and eight other *ESR1* mutations in a total of 329 samples (59-63). The slightly increased rate of D538G mutations compared to other mutations may be a result of the small sample size in our study. We did not detect the K303R mutation in any of our 126 analyzed samples. The prevalence of K303R has been controversial with one group reporting high frequencies of up to 34% (74) and 50% (128) in premalignant and invasive breast cancer respectively, while others have identified it at low frequency (124,129), or not at all (58-63,125-127). The sensitivity of our detection methods suggests that the occurrence of the K303R mutation is likely to be rare.

We detected very high allele frequency (34.3-44.9%) in brain metastases, indicating that the *ESR1* mutant-clones are likely dominant clones, and suggesting that the *ESR1* mutation is a driver event in metastatic progression to this site. Only one bone metastasis had an *ESR1* mutation of relatively low allele frequency (1.4%); however this low frequency is likely due to the very low tumor cellularity in this sample (1-5%). In the future, it might be of interest to test whether different *ESR1* mutations preferentially seed at different metastatic sites.

### **3.4.3 *ESR1* exhibits polyclonal mutations**

Previous studies have shown convergent evolution of polyclonal mutations in cancer, with different mutations in the same gene ultimately targeting the same phenotype (130). We observed cases with multiple *ESR1* mutations in the same tumor, and demonstrated that mutations (Y537S and D538G) were on different alleles, indicating polyclonal disease. Patient CF4 is unique in that cfDNA contained three different *ESR1* mutations. It is possible that the cfDNA integrates *ESR1* mutations from distinct populations of cells, potentially arising from different metastases. The presence of three different mutations in the ligand-binding domain of *ESR1* highlights the substantial selection pressure for these types of mutations during endocrine therapy. Interestingly, longitudinal analysis of cfDNA in this patient indicated increased mutant allele frequency of two clones, and loss of the third clone, possibly reflecting differential response of individual *ESR1* mutations to treatments. There is some prior evidence for different biologies of the different mutants. Toy et al show that ligand-independent activity of Y537S is stronger than that of D538G, and weak for S463P (63). It will be important to investigate if this polyclonality is important in treatment response and tumor progression, e.g. if different clones support each other, or if this simply represents a snapshot of a high rate of genomic instability.

### **3.4.4 Longitudinal monitoring of *ESR1* mutations in cfDNA**

We detected *ESR1* mutations at high mutant allele frequency in cfDNA from patients with advanced breast cancer. The ease of obtaining cfDNA and the high sensitivity suggest that this may be a valuable tool for detecting *ESR1* mutation in patients with advanced breast cancer.

However, larger studies directly comparing *ESR1* mutation in paired cfDNA and metastatic tumor biopsies are required to confirm this possibility. Additionally, cfDNA analysis potentially affords an invaluable approach for longitudinal measurement of mutations that is simply not possible with solid biopsies. This is shown for example in patient CF4 where an initial skin biopsy was negative for *ESR1* mutation, but subsequent cfDNA assays were positive during her course of advanced disease. A study by Mattros-Arruda et al. showed a similar concept in a proof-of-principle study of one patient with advanced disease (68,69). Association between *ESR1* mutation status and response to endocrine therapy is an important question, but our study was not designed to address this. As the numbers were small, retrospective assessment of endocrine therapy history was not examined. Larger studies, and methods to determine the ratio of cfDNA from tumor vs normal cells are required to determine the concordance between primary and metastatic disease, and effects of mutations upon response to hormone therapy. Thus, ultrasensitive detection of rare *ESR1* mutations may represent an important biomarker for development of endocrine resistant disease.

Note: While this manuscript was under review, two other studies reported detection of *ESR1* mutations in cfDNA (131,132).

### **3.5 ACKNOWLEDGEMENT**

This project used the University of Pittsburgh Cancer Institute (UPCI) Biostatistics facility and Tissue and Research Pathology Services that are supported in part by award P30CA047904. The authors acknowledge support of the Health Sciences Tissue Bank staff especially Christina Kline, Louise Mazur, and Christine Thomas. We thank Clinical Research Supervisor Brenda E

Steele and nurses in Magee-Womens Hospital of UPMC for collection of patient blood, Priscilla McAuliffe MD, PhD for assistance with IRB, and many other clinicians and staff, as well as patients at UPMC, for making the study possible. The authors wish to acknowledge the thoughtful input of Dr Daniel M Zuckerman' and the support of Dr Ben Park and David Chu for their invaluable scientific and technical advice.



## **4.0 THE BIOLOGY OF *ESR1* MUTATIONS IN METASTATIC BREAST CANCER\***

**\*Peilu Wang contributed to CRISPR genome editing. Zheqi Li performed Western blots, Adhesion and growth assays. Kevin Levine contributed to the RNA-seq data analysis. Andrew Stern and Zhijie Ding performed ERE-TK assay.**

### **4.1 INTRODUCTION**

As discussed in previous chapters, the *ESR1* gene is significantly mutated in metastatic lesions heavily treated by endocrine treatments. Both tamoxifen and fulvestrant were shown to inhibit the growth of ER mutants, even though higher doses of drugs were required for complete inhibition compared to wild-type ER (60-63). Recent studies have shown that transfection of ER mutant plasmids in cells results in enhanced ligand-independent activity of ER and increased expression of target genes such as TFF1, GREB1 and PR (59-63). Toy *et al* showed that mice bearing tumors with Y537S and D538G mutations grow rapidly in the absence of estrogen (63). A patient derived xenograft (PDX) model of a metastatic breast tumor expressing Y537S also demonstrated ligand independent growth (59). A gene expression analysis of MCF7 cells overexpressing ER WT and mutant constructs indicated constitutive activation of ER target genes in the absence of estrogen. They also identified a subset of novel genes regulated by mutants (63).

Molecular and mechanistic studies have revealed that *ESR1* wild-type and mutants are structurally different. Ser118 is a major phosphorylation site of ER (133). Ser118 is highly phosphorylated in ER mutants, with Y537S mutants showing the highest level (63). Molecular dynamics simulations showed agonistic formation of co-factor bound Y537S caused by the stabilization of H12 helix (61,63). ER mutations are able to confer higher affinity to cofactors such as SRC-1 and SRC-3 (61,63,70,71). Recruitment of cofactors by ER can further promote the regulation of downstream targets required for cell proliferation and survival. Hence, some believe that mutations in ER genes could contribute to major structural changes which lead to ER mediated transcriptome changes.

Taken together, these data demonstrate the constitutive activity of ER mutants in the absence of estrogen or presence of tamoxifen/fulvestrant and suggest a potential role for novel target genes. However, these experiments have all been conducted in transfected models where mutants are expressed many-fold higher than physiological levels. Here in this chapter, we describe the creation of relevant knock-in models by CRISPR technology. Our in-depth RNA-seq analysis demonstrates that *ESR1* mutants have ligand-independent activity of known ER targets, and they also induce a subset of novel genes. Our further experiments show that mutants may drive gain-of-function phenotypes and promote growth, proliferation and survival of the cells in the presence of anti-estrogen drugs.

## 4.2 MATERIALS AND METHODS

### 4.2.1 Cell culture

T47D and MCF7 cells were maintained in RPMI 1640+10% FBS and DMEM+5% FBS, respectively. For estrogen/drug treatment experiments, cells were deprived in IMEM with 10% and 5% CSS for T47D and MCF7, respectively. 17 $\beta$ -estradiol (E2) and 4-hydroxytamoxifen (4OHT) were obtained from Sigma and ICI was purchased from Tocris.

### 4.2.2 Generation of *ESR1* mutant cell lines

The generation of *in vitro* models was performed using CRISPR-Cas9 genome-editing (129,134-137). For the design of the sgRNA, we utilized a web tool (<http://crispr.mit.edu>) entering the sequence flanking the hotspot mutations in the *ESR1* gene (Y537S and D538G) and selected a guide RNA that targets our region of interest (Table 23). The oligos were cloned into PX458 ([www.addgene.com](http://www.addgene.com)), also coding for Cas9, tracrRNA, GFP, and the resulting plasmid was transfected along with the respective double stranded 70bp oligos containing Y537S and D538G mutations into T47D cells. GFP+ cells were sorted via FACS, and the mutation was confirmed by Sanger sequencing (Figure 34). We were able to obtain two clones for Y537S and three clones for D538G mutations along with three WT control clones. We also received MCF7 cells with WT, Y537S and D538G genotypes (two clones each) through collaboration with Dr Ben Park (Johns Hopkins University). They used adeno-associated virus (AAV)-mediated gene targeting as previously published by this group in order to create MCF7 mutants (138).

### **4.2.3 ddPCR**

ddPCR was performed on WT and mutant clones as previously described (67). Briefly, 50-60 ng of DNA was isolated from cells and mixed with ddPCR supermix and Y537S or D538G primer/probe sets. The mastermix was then partitioned into droplets by Biorad droplet generator. The PCR was conducted subsequently using 40 cycles of amplification. The fluorescent signal from each droplet was finally quantified by Biorad QX100 system and analyzed by QuantaSoft software.

### **4.2.4 Western blot**

MCF-7 and T47D clones were counted after 3 days of hormone deprivation in CSS, and plated into 6 well plates with the concentration of 120,000 cells per well (MCF-7) and 90,000 cells per well (T47D). Cells were treated with 0.1% ethanol as a vehicle control or 1 nM of E2. The cells lysed with RIPA buffer and subsequently sonicated. Protein concentrations were determined with BCA assay kit following the manufactural protocol (Thermo Fisher Scientific). 80 ug of proteins per samples were loaded in SDS-PAGE gel, and then transferred onto PVDF membrane. The antibodies against total estrogen receptor and phosphor-estrogen receptor (Ser118) were purchased from Cell Signaling Technology and Signalway Antibody, respectively.

### **4.2.5 Transcriptional reporter activity of WT and mutant *ESR1*\***

We set up a 384-well plate format using the ThermoScientific Multidrop Combi for cell dispensing, Velocity 11-Bravo liquid handling instrument for compound dispensing, and the

Perkin Elmer EnVision multilabel reader for luminescence detection resulting from the ERE-Tk-luc reporter (63). The pRL-TK renilla reporter was used to normalize transfection conducted with X-tremeGene HP transfection reagent. The cells were stimulated by increasing doses of E2 for 24 hours and luminescence was read accordingly

\*(performed by Dr Andrew Stern and Dr Zhijie Ding at the Drug Discovery Institute, University of Pittsburgh).

#### **4.2.6 RNA-seq analysis**

Each clone for WT and mutant T47D and MCF7 cells were deprived of estrogen for three days. The clones for each genotype (WT or muts) were then pooled and plated in four replicates in 6-well plates. The cells were treated by veh or E2 for 24 hours and RNA was isolated via Qiagen RNeasy kit according to the manufacturer protocol.

500ng RNA of each sample was sent to the genomic core at Children's Hospital of Pittsburgh and subjected to NGS obtaining >15M reads per sample. We used Salmon for quantification of the transcripts using default options and hg38 genome build as the reference (139). The differentially expressed (DE) genes between WT and mutants were identified by DEseq2 package using the contrast option (140). The genes which had a maximum transcripts per million (TPM) <1 were excluded from the analysis due to avoid confounding pathway analysis with genes of low expression. We employed R to plot the heatmaps and perform the statistical analysis. To test the statistical significance of overlapped genes in venn diagrams, we used Chi-square test.

#### **4.2.7 Growth Assay**

MCF-7 or T47D cells were evenly pooled after 3 days of hormone deprivation in CSS, and plated into 96-well plates with concentration of 2500 cells per well (MCF-7) or 4000 cells per well (T47D). After 24 hours, the cells were treated with 0.1% ethanol as vehicle control, 1 nM of E2, 100 nM of fulvestrant or their combination. The cells were harvested for quantification after 0, 2, 4, 6 and 9 days with the FluoReporter kit (Life Technology) following the manufacturer's protocol. The data was plotted and IC-50 was calculated by PRISM statistical package.

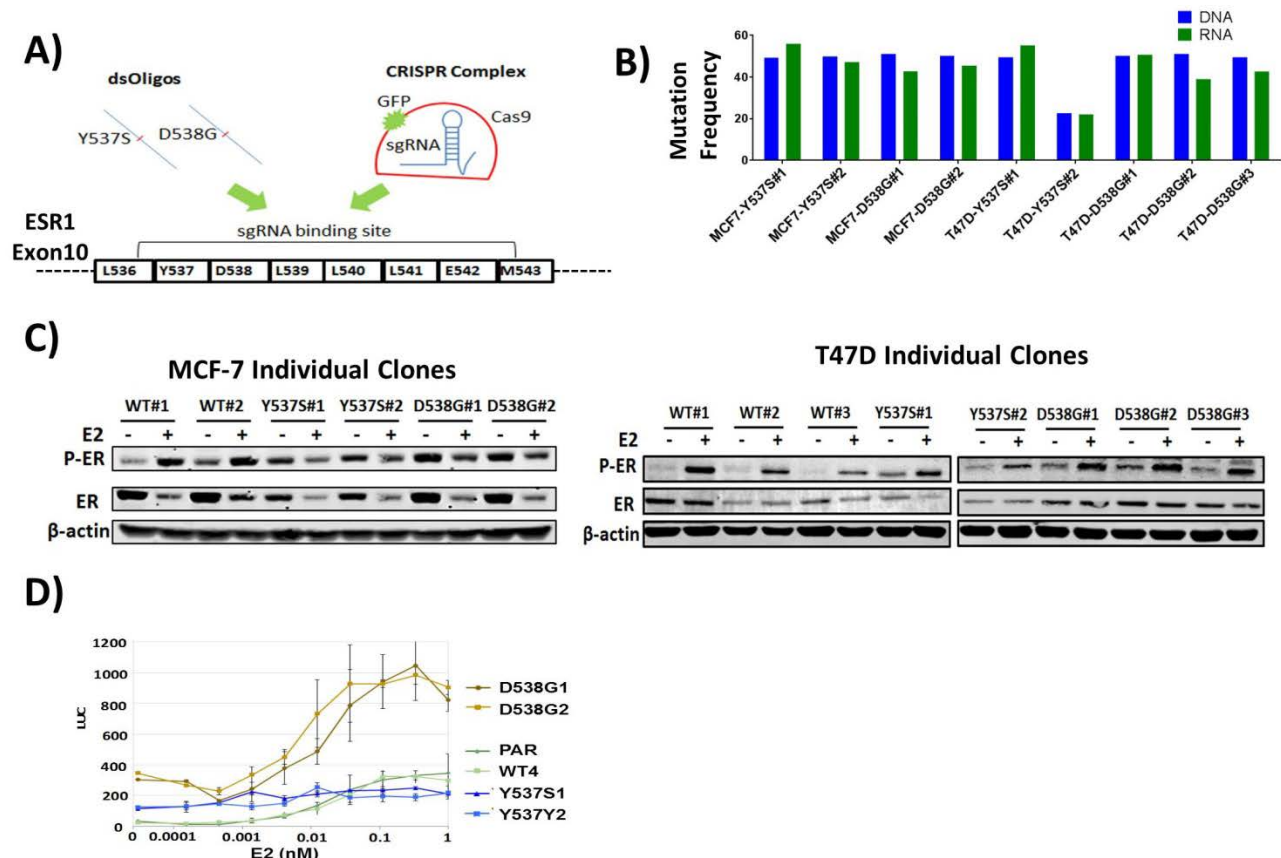
#### **4.2.8 Adhesion Assay**

MCF-7 or T47D cells were counted first and were then added into Collagen I (Thermo Fisher Scientific), Collagen IV (Corning)- or BSA-coated 96-well plate as well as uncoated 96-well plates with the concentration of 30, 000 cells per well. For the ECM array assay, T47D cells were resuspended in serum-free medium first, and added into the ECM array plate (EMD Millipore) with 100, 000 cells per well. After incubation in 37 degree for 2 hours, all the coated plates were washed with DPBS for three times. The quantification of cell numbers was performed using the FluoReporter kit (Life Technology) following the manufactural protocol on all the washed-coated plates and the unwashed-uncoated plate. The adhesion ratio was calculated by dividing the values of washed to unwashed wells.

## 4.3 RESULTS

### 4.3.1 Molecular characterization of *ESR1* mutations

To study the biology of *ESR1* mutations in breast cancer, we employed two ER-positive cell lines that are well-studied in the literature. T47D cells were genome edited via CRISPR technology resulting in two and three clones of Y537S and D538G, respectively (Figure 11A). We also received MCF7 cells harboring Y537S and D538G mutations (two clones each) from Dr Ben Park Lab. The mutation allele frequency of each clone was inspected in DNA and RNA by highly quantitative ddPCR using corresponding probe/primer sets (Figure 11B). The ER mutations frequency was approximately 50% showing heterozygous loci in all clones except T47D Y537S#2 clone where the mutation is only expressed in 22% of DNA and RNA. This could be due to more copies of WT ER copies in this clone.



**Figure 11. Generation and molecular characteristics of *ESR1* mutations.**

A) A schematic view of CRISPR genome-editing: Cas9 and sequence specific sgRNA constructs from the CRISPR complex which makes a cleavage in the target DNA. Homologous recombination pathway in the cells repairs the DNA using template oligos which contain mutations of interest. B) The mutation frequency of each clone was checked by ddPCR at the DNA and RNA level. C) Western blotting of mutant and WT clones of T47D and MCF7 cells. The blots were stained for ER and Phospho-ER (P-ER). Clones for WT or mutants were pooled and estrogen deprived for three days. Cells were then treated by Veh or E2 (1nM) for 24 hours. D) Transcriptional activity of WT and mutant cells were analyzed by high-throughput ERE-TK assay using different doses of E2 in T47D cells following three days of E2 deprivation. (C provided by Zheqi Li, D provided by Dr Andrew Stern and Zhijie Ding)

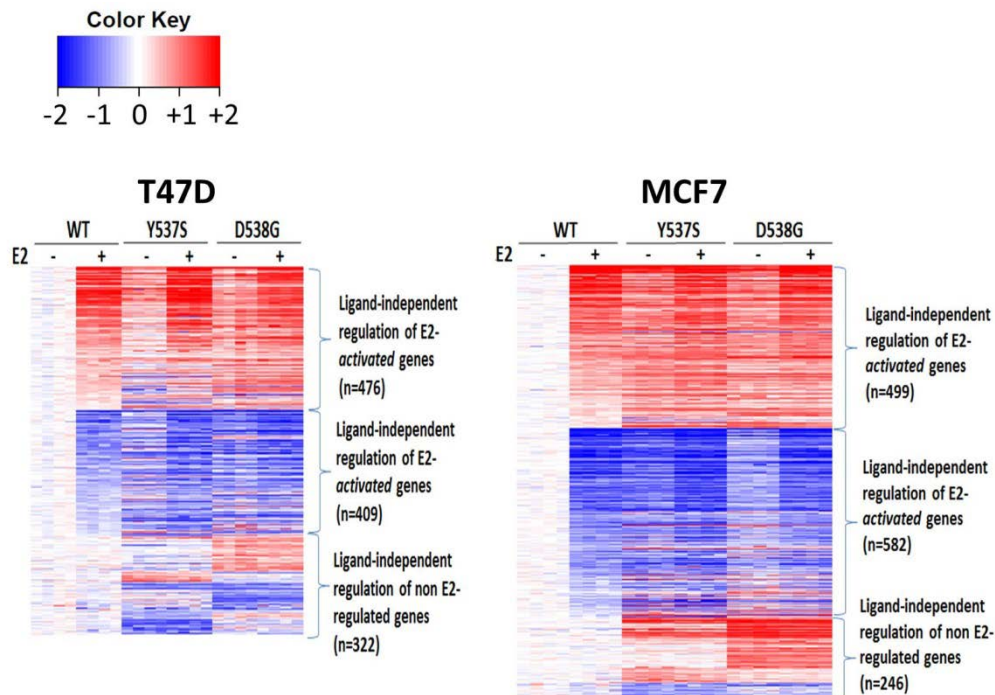
We next tested the protein expression and phosphorylation of ER in all clones (Figure 11C). More variation was observed in baseline ER levels among T47D clones compared to those of MCF7. Phosphorylation at Ser118 has been shown to be required for the full activity of ER (141). We observed that mutants displayed higher constitutive ER phosphorylation compared to WT in both cell lines. The level of phosphorylation decreased after E2 treatment in MCF7 mutants while it increased in T47D mutants. This suggests cell line specific ER activity that might be dependent on the unique molecular profile of each model (141).



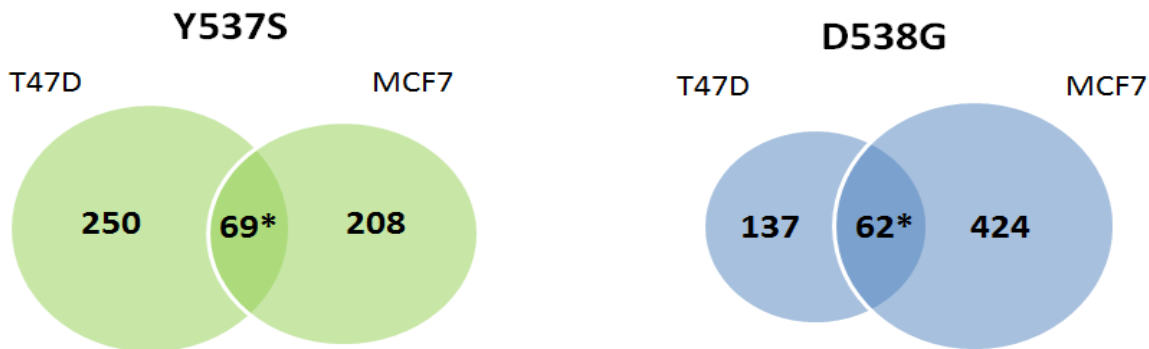
To test transcriptional activity of the ER mutant proteins, we set up a high throughput 384-well plate ERE-tk-luc reporter assay system. We detected ligand-independent activity of the ER mutants, and in addition, D538G was hyper-responsive to estradiol in T47D cells (Figure 11D). This data suggests that both mutations are active in the absence of estrogen and further, D538G could induce/repress target genes at lower doses of E2.

#### **4.3.2 Transcriptome regulation by ER mutants**

To further assess how mutations could impact the transcriptional activity of ER, we performed whole transcriptome RNA-seq of the mutant cells in the absence and presence of E2. PCA analysis of the top 1,000 most variable genes showed that biological replicates cluster together (Figure 35). The gene expression analysis characterized a total of 1,327 genes for Y537S and 1,207 genes for D538G that were differentially regulated compared to WT cells in the absence of ligand (cutoff=fold change >2, p-value<0.005) (Figure 12). The heatmaps in Figure 12 show that approximately 2/3 of these genes were estrogen regulated in WT clones, suggesting that mutants are more active than WT in the absence of E2. This was consistent with our previous transactivation data showing the constitutive activity of mutants. We also observed a high overlap in constitutively regulated genes between the two cell lines (chi-square test, p-value<0.01) (Figure 13)



**Figure 12. Heatmap of ligand independent differentially regulated genes between WT and mutants.** Individual clones were estrogen deprived for 3 days and pooled for each genotype (WT, Y537S and D538G). Cells were then treated by Veh or E2 (1nM) for 24 hours. RNA-seq was performed on RNA isolated from the cells and differentially expressed genes were called by DESeq2 (fold>2, adjp<0.005). Color represent log2 fold change.  
(In collaboration with Kevin Levine)



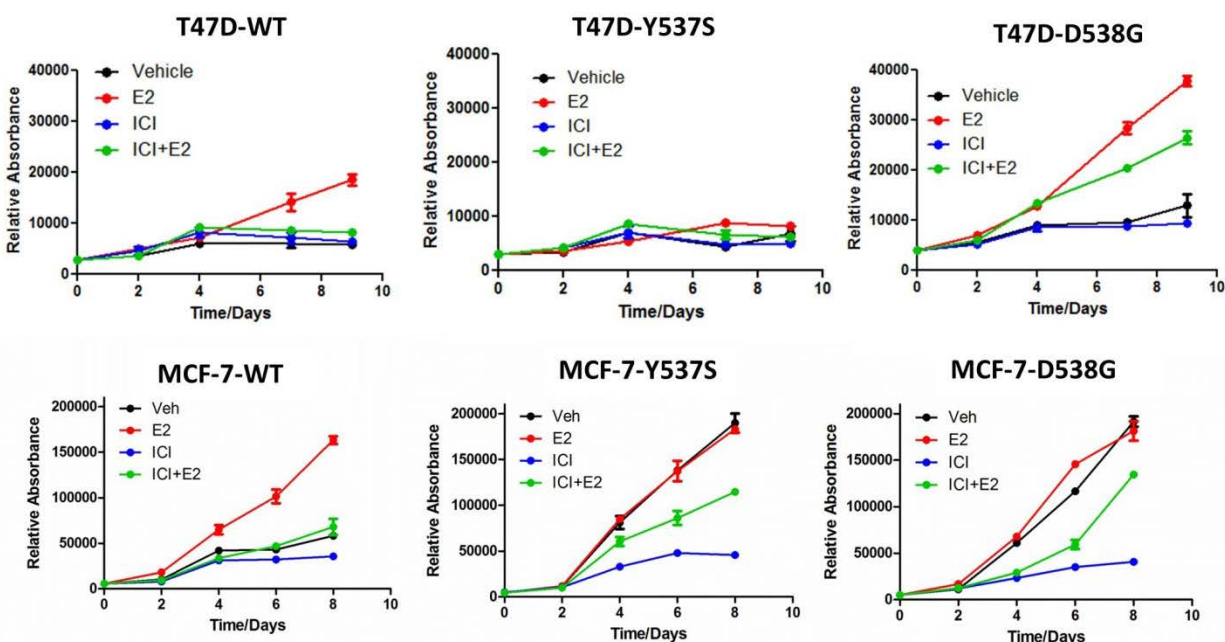
**Figure 13. The overlap of ligand independent regulated genes between the cell lines for each mutations.** The ligand independent genes (p-value<0.005, fold change>2) were overlapped between the two cell lines. The Chi-square test was used to test whether the overlap is significant.  
\*p-value<0.01

Most importantly, however, a set of genes (n=246 in MCF7 and n=322 in T47D) were uniquely regulated only in mutant clones. In T47D cells, 304 and 187 genes were differentially expressed in Y537S and D538G, respectively, with 40 genes overlapped between the two mutations (chi-square test for overlap, p-value<0.01). In MCF7 cells, 241 and 244 genes were induced or repressed in Y537S and D538G, respectively, with a larger statistically significant overlap of 105 genes (chi-square test, p-value<0.01) (Figure 36). We also found that only 12 genes were shared between T47D and MCF7 mutants (11 in Y537S and 1 in D538G) among the ligand independent potential novel targets of ER (Table 24), although this did not reach statistical significance. This result suggests that the gain-of-function of properties of the mutations may be cell line specific and dependent on the genetic background of knock-in cells.

IPA pathway analysis showed enrichment of cancer and immunological diseases in mutants of both cell lines (p-value<0.001). Given the fact that there was little overlap of novel genes between the two cell lines, we performed the pathway analysis on the novel regulated genes shared between Y537S and D538G in each cell line independently (Figure 39). We found that several upstream regulators are activated in mutants such as STAT and Interferon signaling in MCF7, and FOS, TGFB1 and SMAD4 in T47D mutants. The upregulation of these pathways have been previously shown to be associated with breast and other types of cancer (26,142-152). Taken together, our data suggests that mutants could potentially activate alternative pathways in addition to ligand-independent classic ER pathways.

### 4.3.3 Gain of function activities of *ESR1* mutants in genome edited cell lines

Previous overexpression models of mutant ER have shown ligand-independent growth as well as partial resistance to antiestrogens (60-63). We investigated the ligand-independent growth of our knock-in models and found that T47D-D538G cells proliferated at a higher rate in the absence of E2 and are hyper sensitive to E2 compared to WT clones. T47D-Y537S cells, however, did not show a similar behavior. In the MCF7 cell line, both mutants displayed higher ligand independent growth and were at least as responsive to E2 as WT cells (Figure 14).



**Figure 14. Growth assay in WT and mutant cells.**

T47D and MCF7 cells were deprived for 3 days followed by different treatments: E2 (1nM), ICI (100 nM), and E2 (1nM) + ICI (100nM) (Provided by Zheqi Li)

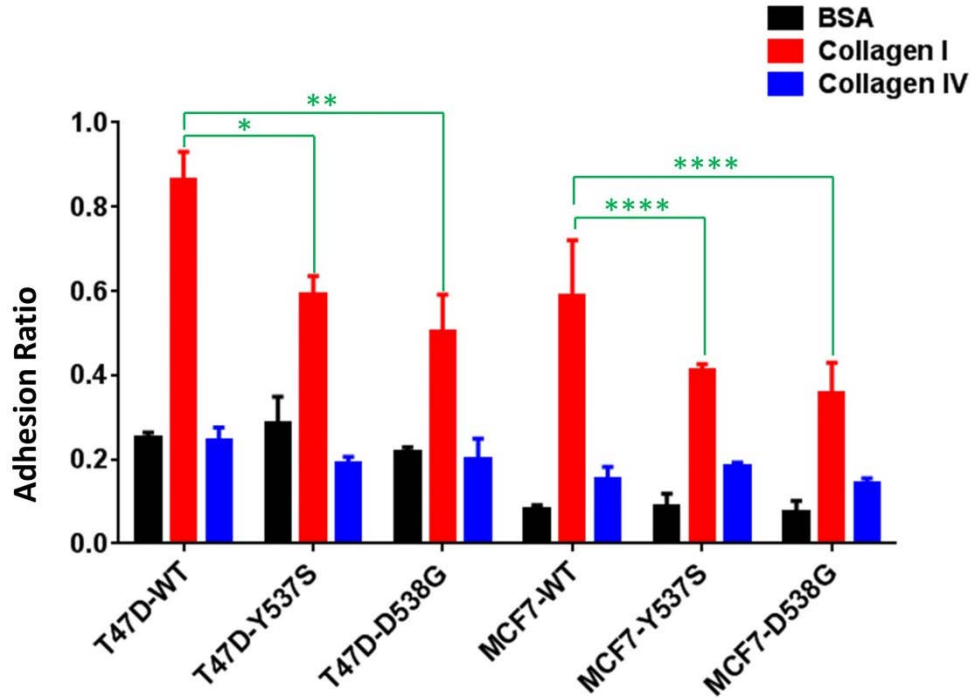
To test the response of mutants to anti-ER drugs, we performed a growth study of cells treated by various doses of ICI, Raloxifene and 4OHT. The  $IC_{50}$  for ICI was significantly higher in MCF7 mutants and to a lesser degree in T47D mutants compared to WT (Table 7). In both cell lines, Y537S mutations represented a remarkable resistance to 4OHT and Raloxifene compared

to D538G and WT clones. This suggests that the Y537S mutation may have a survival advantage in tumors treated by SERMs.

**Table 7. The IC-50 for WT and mutant cells treated by different compounds**

Compound	T47D-WT	T47D-Y537S	T47D-D538G	MCF-7-WT	MCF-7-Y537S	MCF-7-D538G
ICI	0.42 nM	1.41 nM	1.32 nM	0.72 nM	30.57 nM	4.49 nM
4OHT	0.27 nM	11.66 nM	1.33 nM	0.49 nM	211.15 nM	1.56 nM
Raloxifene	0.09 nM	2.47 nM	0.24 nM	1.48 nM	51.44 nM	11.19 nM

Tumor metastasis involves adhesion molecules that are responsible for the attachment and detachment of cells (153,154). Several studies have shown a role for Collagen I and IV in promoting cancer metastasis (155-158). We examined the adhesion of WT and mutant cells (pooled clones for each genotype) on plates coated with Collagen I and IV (Figure 15). Interestingly, both D538G and Y537S cells showed less adhesion to Collagen I (p-value<0.05 in T47D and p-value<0.0001 in MCF7) whereas attachment to Collagen IV was not different from WT cells. We performed the assay on individual clones and observed consistent results (Figure 37). A similar experiment was performed on an array of extracellular matrices (ECMs) using T47D clones and observed less adhesion to Collagen II, Fibronectin, Laminin, Tenascin and Vitroectin in D538G or both mutants compared to WT (Figure 38). These data further indicate that less adhesion of mutants to basement proteins may confer a metastatic phenotype to cells.



**Figure 15. Cell adhesion to Collagen I and IV.**

The individual clones of each genotype (WT and mutants) were pooled and plated on precoated plates with BSA (negative ctrl), Collagen I and Collagen IV. The statistical significance of difference was analyzed by ANOVA test corrected for multiple comparison (\*p-value<0.05, \*\*p-value<0.01, \*\*\*\*p-value<0.0001). (Provided by Zheqi Li)

#### 4.4 DISCUSSION

In this study, we report for the first time the generation of knock-in models of *ESR1* mutations, Y537S and D538G. Previous studies have all employed cell lines transfected with constructs expressing significantly high levels of mutations which do not recapitulate what has been observed in actual tumors (57,59-63). Here we used CRISPR technology to genome-edit T47D cells that resulted in at least two clones for each mutation. As controls, we selected three WT clones which underwent CRISPR transfection, but maintained an intact *ESR1* locus. To avoid potential off target activity, we designed a unique guide RNA based on previously described methods (135,137,159). In addition, we generated more than one clone for each mutation and

used two independent cell lines to minimize the possibility of data confounded by clone specific features.

We next characterized the mutation frequency, ER expression and phosphorylation in all clones. T47D-Y537S clone #2 represented a lower mutation allele frequency compared to other mutants suggesting that there are more WT copies of *ESR1* gene present in this clone (Figure 11B). However, this outcome is not concerning in our study since previous reports show variable *ESR1* mutant allele frequency (mainly <50%) in metastatic lesions (60,62,63). We found that total ER levels were different between individual clones, which may be due to clonal variation during the CRISPR selection. The mutants consistently showed higher phosphorylation of ER in the absence of estrogen (Figure 11C). This is in line with previously published data, although the increased phosphorylation of ER was not as high as previously reported (63). This could be due to significantly high expression of mutations in transfected models. Phosphorylation of ER at this particular site was previously reported to be elevated in tumor biopsies taken from patients who had relapsed following tamoxifen treatment (160). Constitutive activity of ER mutants was further confirmed in T47D cells by ERE-tk-luc in which both Y537S and D538G showed transcriptional activity in the absence of E2.

Transcriptomic analysis of *ESR1* mutants was performed using RNA-seq in cells treated with veh or E2. We identified a large subset of genes that were constitutively regulated in mutants and regulated by E2 in WT cells. These mostly include classic ER target genes required for cell growth and proliferation such as IGFBP4, GREB1, MYC, and TFF1. Intriguingly, we found a number of genes that are not E2 regulated in WT but are uniquely induced/repressed in mutants. Pathway analysis revealed that several cancer associated networks are activated by

these novel genes. We are performing additional experiments to characterize the specificity of these pathways in our mutant models.

In one study, the gene expression of mutants was investigated in MCF7 cells transfected with the various ER $\alpha$  mutant constructs in hormone-depleted medium (63). The investigators found a panel of 92 genes differentially regulated in mutants vs WT in MCF7 cells. This data is, however, subjected to some bias as the mutation levels were non-physiological due to the nature of overexpression models. Furthermore, the E2 regulated genes in WT were not subtracted in their data.

Previous studies have shown that *ESR1* mutations occur in disease refractory to endocrine treatments (59-63). Limited data from these studies on overexpression models have suggested partial resistance to SERMs and SERDs with the latter being more potent in inhibiting ER mutations. We tested the proliferation of mutants exposed to different compounds including E2, 4OHT, raloxifene and ICI in a dose-dependent manner. In addition to constitutive activity of both mutations, we observed hyper-sensitivity of D538G in T47D cells. Both mutants also showed resistance to physiological levels of anti-ER drugs which were able to fully inhibit WT clones. Y537S, however, showed a higher resistance to SERMs compared to D538G. This data may explain the higher frequency of D538G in metastatic lesions where patients were more likely to be treated by AIs shown in previous clinical studies (60-63,65). Future prospective clinical studies are required to discover the drug specific characteristics of the mutations.

Adhesion molecules have been long known to play a critical role in the process of cancer metastasis (161-163). Our adhesion assay showed that mutants are significantly less attached on Collagen I coated basement. It has been previously indicated that highly metastatic cells express a decreased binding to Collagen I, which results in more motility and invasiveness (164,165).



Experiments are ongoing to investigate the metastatic phenotype of our mutant models through motility and migration assays.

## 5.0 CONCLUSIONS

### 5.1 IDENTIFICATION AND FUNCTIONAL ASSESSMENT OF ER REGULATED SNVS IN BREAST CANCER

Breast cancer is the most common type of cancer in females and the second leading cause of death due to cancer. The majority of breast tumors overexpress ER, which is known to be a driver of cancer in these tumors. As we discussed in earlier chapters, the DNA binding profile of ER varies among tumors and may be associated with prognosis and response to endocrine treatment. However, genomic alterations underlying differential ER binding to DNA remains understudied.

In this dissertation we addressed the above question by developing a pipeline to identify potential regSNVs in ER ChIP-seq studies performed on breast cancer models. We first tested our pipeline on ER binding sites extracted from studies of the MCF7 cell line and identified a functional SNV in the IGF1R gene. The minor allele rs62022087 was predicted in our analysis to strengthen an ER binding site in the second intron of the IGF1R gene. Further *in vitro* experiments confirmed and validated that this regSNV is able to increase ER affinity to DNA and therefore, increase the expression of IGF1R gene.

We next applied our pipeline to all available ER ChIP-seq data sets in breast cancer. This analysis led to the discovery of a large number of potential regSNVs enriched in intergenic and

intronic sites. By integrating TCGA data, we found that regSNVs are enriched in proximity of genes differentially regulated between ER+ and ER- disease. Further, we asked whether any of the regSNVs were associated with the expression of corresponding target genes. We identified 17 variants regulating the expression of host genes by modifying ER binding. Our top candidate was rs36208869 in the promoter of GSTM1 gene whose expression was correlated with survival in breast cancer patients. Future studies are required to study and characterize the clinical significance of RegSNVs in breast cancer.

In the last part of chapter 2, we searched for somatic mutations in ER binding sites by integrating WGS and ChIP-seq data. Our initial analysis of published WGS on 45 tumors identified a panel of recurrent somatic regSNVs located in potential ER binding sites. Two mutation hotspots were found in the intronic regions of GPR126 and PLEKHS1 genes which were further confirmed in a cohort of 98 breast tumors in TCGA. Several reports have indicated the PLEKHS1 mutations in cancer (81,82,105) but mutations in GPR126 were not identified previously. We did not see any correlation between the presence of mutations and expression of target genes in TCGA. In collaboration with Dr Geoff Greene, we are performing experiments similar to 4C-seq in T47D cells in order to identify long-range targets of these potential regulatory elements. We will validate the identified targets in additional cell lines and tumor samples which will be eventually tested whether they may impact the clinical outcome in breast cancer patients.

One of the limitations in our study is using various types of data generated by different platforms and labs. This may introduce some error into our analysis due to technical variation. Another challenge which we faced in our study was lack of normal samples in ChIP-seq data which made it difficult to differentiate between germline and somatic variants. We used dbSNP

and 1000Genome databases to annotate the variants which have been reported to be germline. In order to systematically study regulatory SNVs in breast cancer, we need to conduct ChIP-seq and multi-omics analysis on the tumor sample and obtain clinical information by following up patients for long term outcome.

The role of non-coding genomic variants in cancer and other diseases has been largely understudied due to the technological challenges and lack of understanding about the non-coding genome. In this dissertation, we present a novel pipeline to capture regulatory SNVs by integrating multi-omics data and validate them through *in vitro* studies. We believe our methodology is applicable to not only other types of cancer, but also other genetic based diseases. The screen for impactful regulatory variants will soon become part of genetic testing as our knowledge of non-coding genome improves and sequencing hurdles are being lifted. Such genetic tests are of great importance to public health in order to tailor the treatment to the needs of each individual patient. Public health authorities will be able to treat cancer patients in a personalized manner and cooperate with insurance companies to cover more targeted therapies. This will eventually reduce the cost burdens on public health and direct the patients towards the appropriate treatments.

## **5.2 DETECTION OF *ESR1* MUTATIONS IN PRIMARY TUMORS, METASTATIC LESIONS AND CFDNA OF PATIENTS WITH ADVANCED BREAST CANCER**

As mentioned in chapter 3, multiple studies have shown a high frequency of *ESR1* mutations in metastatic breast cancer (59-63). However, technical limitations hindered scientists from studying rare *ESR1* mutations in primary tumors and circulating DNA from metastatic patients.

Moreover, the frequency of ER mutations was not well studied in lesions hard to biopsy such as bone and brain. This prompted us to use the highly sensitive ddPCR technology to address these issues.

Six *ESR1* mutations were assessed in clinical samples from a total of 121 patients. Mutation rates were 7.0% (3/43 primary tumors), 9.1% (1/11 bone metastases), 12.5% (3/24 brain metastases), and 24.1% (7/29 cfDNA). Two patients showed polyclonal disease with more than one *ESR1* mutation. Mutation allele frequencies were 0.07% to 0.2% in primary tumors, 1.4% in bone metastases, 34.3 to 44.9% in brain metastases, and 0.2% to 13.7% in cfDNA. In cases with both cfDNA and metastatic samples (n=5), mutations were detected in both (n=3) or in cfDNA only (n=2). Treatment was associated with changes in *ESR1* mutation detection and allele frequency. Low allele frequency in some primary tumors suggests that in some tumors, rare *ESR1* mutant clones are enriched by endocrine therapy. Further studies should address if sensitive detection of *ESR1* mutations in primary breast cancer and in serial blood draws is predictive for development of resistant disease.

It has been shown that *ESR1* mutations in metastases may contribute to a shorter progression free survival (65). Several other studies have shown that these mutations could drive metastasis in patients treated with endocrine therapy (discussed in chapter 3). Therefore, it is very important to screen the patient with ER+ disease on a regular basis for early detection of ER mutations. In this dissertation, we describe a highly sensitive tool to achieve this goal. Using ddPCR technology, we were able to detect low frequency *ESR1* mutations in the blood of the patients with advanced disease. From the public health perspective, genomic testing via cfDNA is preferred by patients and clinicians over traditional invasive means such as surgical biopsies. Furthermore, utilizing liquid biopsies for monitoring cancer treatment and progression will

decrease the need for costly imaging tools and therefore, reduce the costs on public health agencies. Clinical studies are ongoing to validate the sensitivity and specificity of liquid biopsies including cfDNA in different types of cancer to treat patients more effectively in the future.

### **5.3 FUNCTIONAL ANALYSIS OF *ESR1* MUTATIONS IN ENDOCRINE TREATMENT RESISTANCE BREAST CANCER**

Point mutations in the *ESR1* gene are clustered in the LBD domain of ER protein, suggesting a gain of function for resistance to anti estrogen treatments (discussed in chapter 1 and 4). The biology of ER mutations has been studied mainly in transfected models where the mutants are expressed at very high levels that are not physiologically relevant. In order to create appropriate knock-in models, we used CRISPR technology to genome-edit T47D cells. We were able to obtain several clones for each of the most frequent ER mutations, Y537S and D538G. In addition, MCF7 knock-in clones created by Dr Park's Lab were included in our functional studies.

The genotype of each clone was first confirmed by Sanger sequencing and ddPCR at the DNA and RNA level. All the clones expressed ER protein although the levels were variable in T47D cells. ER phosphorylation at Ser118, a marker of transcriptional activity, was higher in mutant clones, and this was consistent between T47D and MCF7 cell lines. ERE-TK transactivation assays showed that mutants can constitutively activate transcription of luciferase with D538G being more responsive to lower doses of E2.

We next sought to characterize the transcriptomic changes mediated by ER mutants. The majority of the genes differentially regulated between WT and mutants at baseline were

induced/repressed by E2 in WT. Furthermore, a subset of genes were uniquely regulated only in mutant clones and speculated to be novel targets of ER mutants. Pathway analysis showed that these novel targets can activate metastasis associated pathways such as STAT, Interferon, SMAD4, and TGFB1. Our future plan is to target ESR1 mutants by high doses of SERDs such as ICI and see if the effect of novel targets could be reversed. Furthermore, we are performing ChIP-seq experiments in mutant cells to identify potential novel binding sites of ESR1 mutants which may drive their gain of function in the cell.

As we discussed in previous chapters, *ESR1* mutations are selected as a result of anti-E2/ER treatments. We measured the growth of WT and mutant cells in the presence of 4OHT, Raloxifene and ICI. Both mutants demonstrated partial resistance to drugs with Y537S being less sensitive to SERMs. This confirms the clinical relevance of mutations in the endocrine treatments settings. Finally, we investigated cell adhesion as a metastasis associated phenotype in WT vs mutant cells. Interestingly, both mutants in T47D and MCF7 cells showed less attachment to Collagen I. It is plausible that mutants have gained properties to digest Collagen I or similar scaffolds and become more invasive. Our studies are now ongoing to assess the motility and aggressiveness of ER mutants on Collagen I precoated plates.

We faced a few challenges in the study of ESR1 mutations. Single mutant cells were selected in the process of CRISPR genome editing and this may introduce clonal bias into our downstream functional experiments. To address this concern, we created more than one clone for each mutation and used all clones in our validation experiments. However, we were obliged to pool the clones of each genotype in some other experiments due to lack of resources such as RNA-seq. Another challenge in our study was to select the appropriate breast cancer cell lines for mutating ER. We genome-edited the T47D cell line via

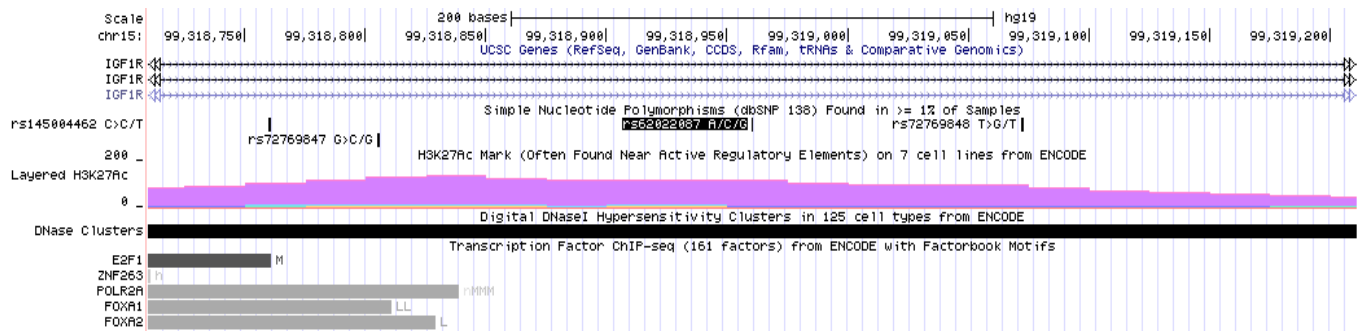
CRISPR and also received MCF7 mutants from Dr Park's lab. Both of these cell lines are well known to be estrogen responsive and may be good models as recipients of ER mutations. However, the genetic and proteomic background of T47D and MCF7 cells are different, which can lead to differences in endpoint phenotypes. This may explain why the same mutations exhibit functional differences in some of our assays between the two cell lines.

Understanding the biology of ER mutations is critical for targeting the mutant cells in the tumor more efficiently. This dissertation included some preliminary data on the biology of ER mutations in more appropriate models compared to previous studies. Partial resistance to SERMs and SERDs in mutants suggests that higher doses of these drugs are needed when ER mutations are present, although this needs to be tested in future clinical trials. Our data also suggests some gain-of-function activity at the transcriptomic levels which may lead to novel phenotypes in mutants. This may be mediated via novel interaction with cofactors driving ER transcriptional complex to new targets. We are implementing new studies to test these hypotheses and target not only ER mutations, but also alternative networks cooperating with ER mutant protein.

Endocrine treatment in breast cancer has public health implications given the fact that it is the most successful targeted therapy in ER+ disease, which comprises 70-80% of all breast cancers. A large proportion of patients, however, acquire resistance during or after the course of treatment due to ESR1 mutations. Costs associated with illness and therapies in metastatic breast cancer patients are remarkably high. Therefore, public health authorities need to allocate sufficient budgets for finding novel therapies against ER mutants that can significantly increase survival and quality of life in metastatic patients.



## APPENDIX A: SUPPLEMENTARY FIGURES



**Figure 16. The UCSC genome browser view of the second intron in IGF1R gene.**

The index SNP, rs62022087, seems to be located in a region bound by several chromatin modifying factors based on ENCODE data.

A)

MCF7\_1 (100%)

MCF7\_2 (100%)

MCF7\_3 (100%)

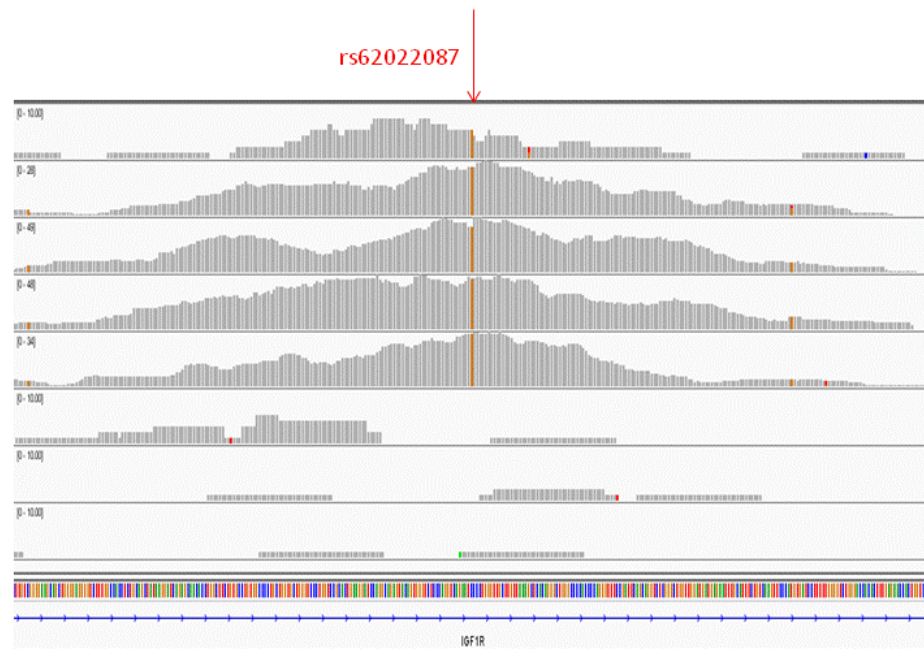
MCF7\_4 (100%)

MCF7\_5 (100%)

T74D\_1 (0%)

T47D\_2 (0%)

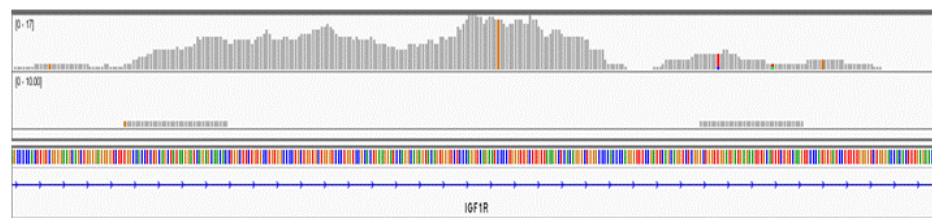
ZR75\_1 (0%)



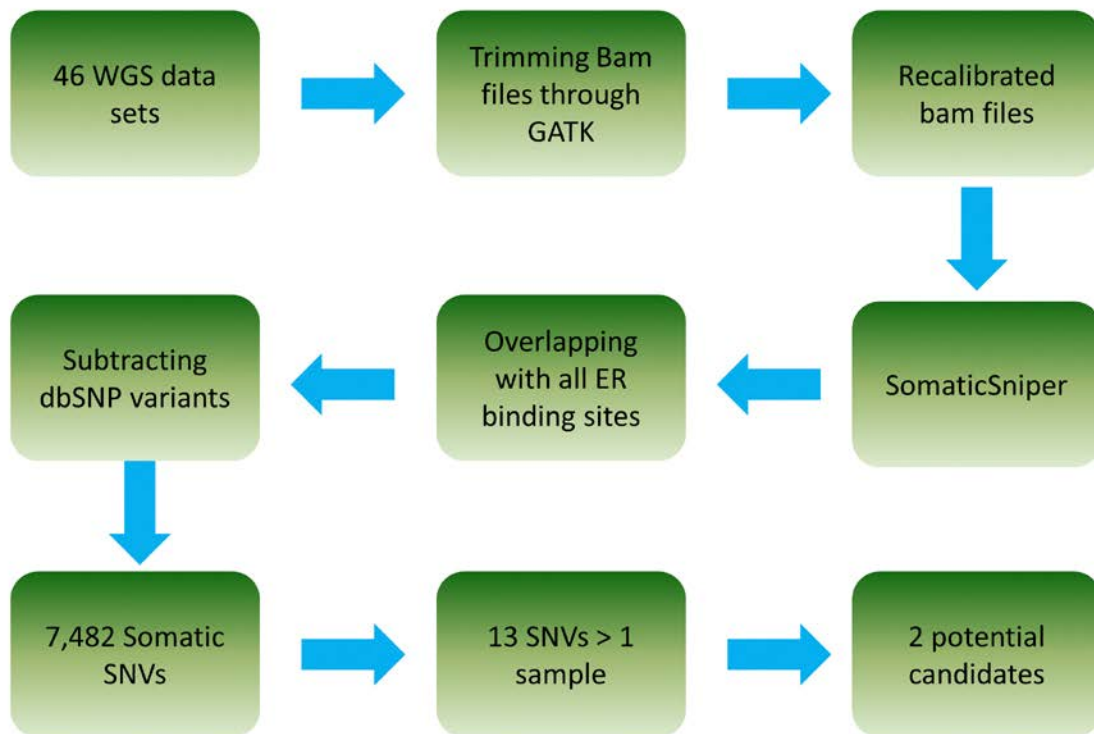
B)

MCF7 (100%)

T47D (0%)

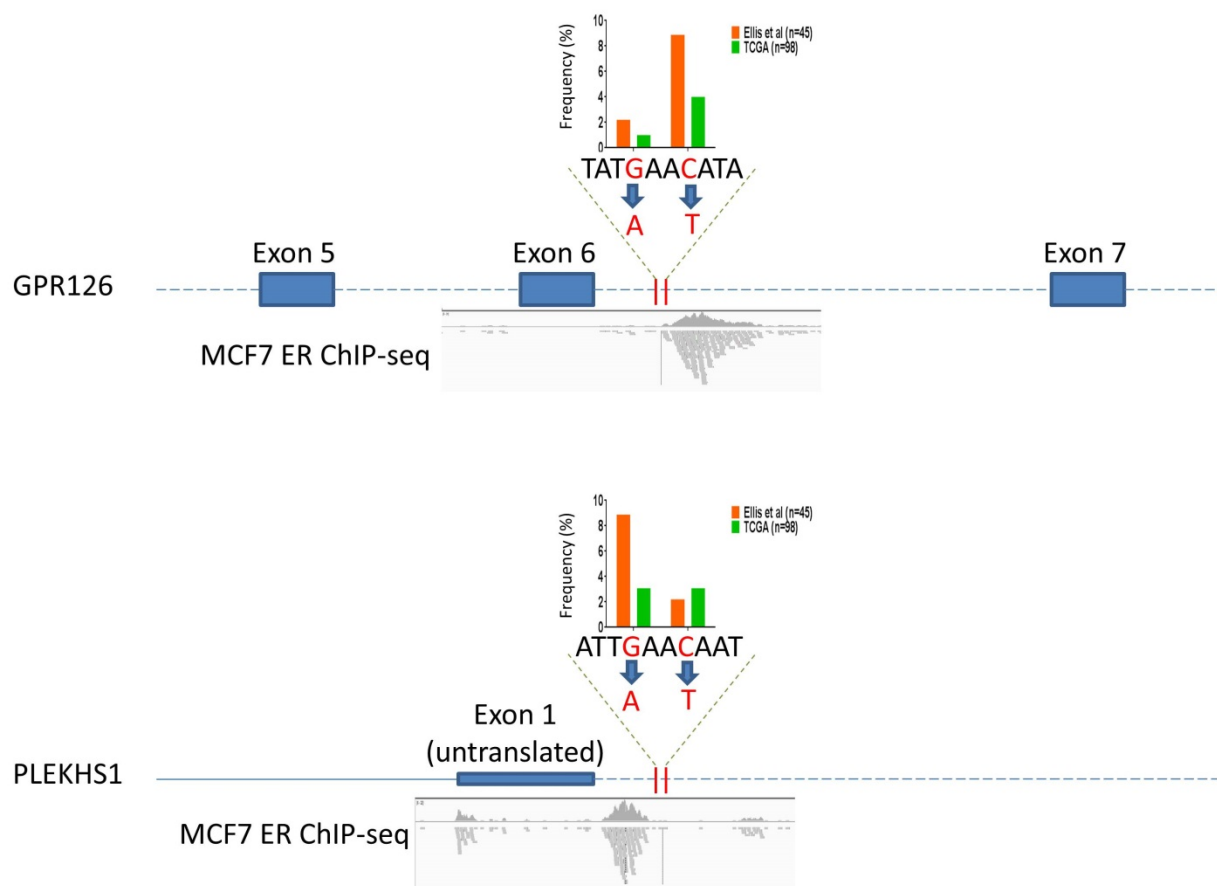


**Figure 17. The visualization of ChIP-seq reads from multiple cell lines over rs62022087 SNP site.**  
Data extracted from Ref: (3,44)



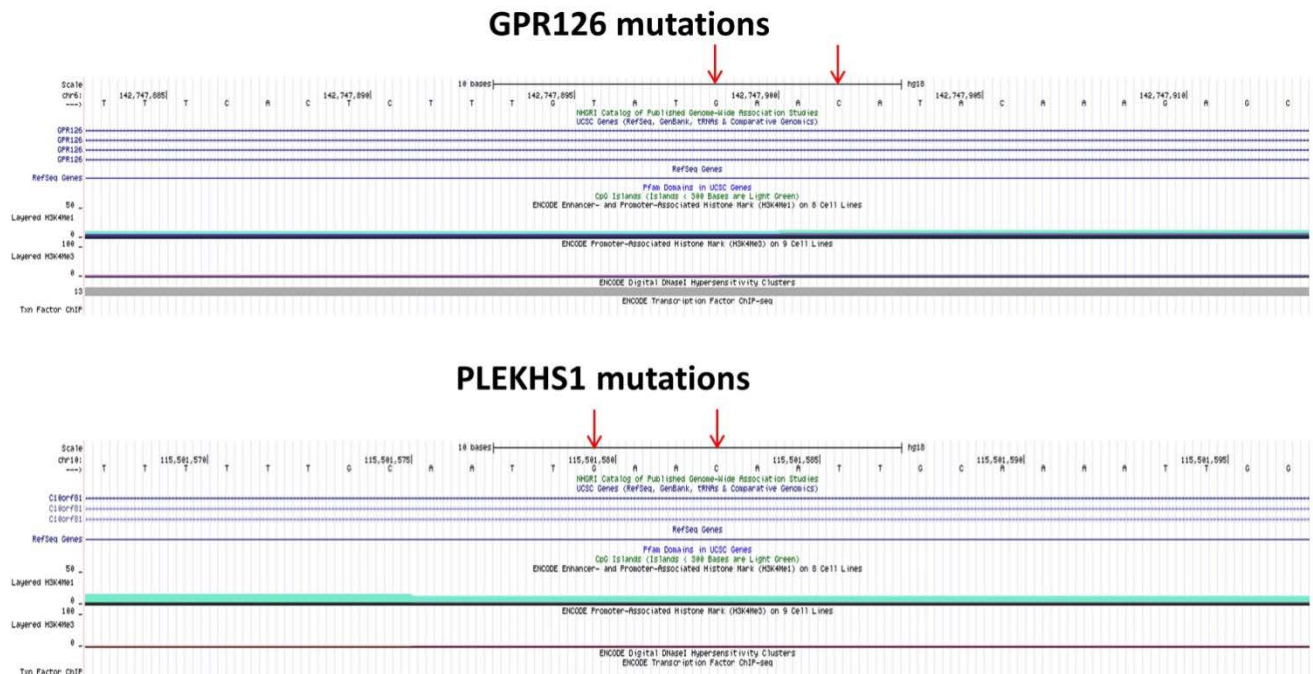
**Figure 18. The pipeline used for analyzing breast tumors and calling somatic SNVs within ER binding sites.**

GATK was used to refine the raw bam files. SomaticSniper was further utilized to call the somatic mutations from the bam files. To further filter out the potential contamination from normal samples, we subtracted dbSNP variant from our list.



**Figure 19. Schematic view of somatic non-coding mutations found in two independent WGS data sets.**

The bar graphs show the frequency of the mutations in each data set. The MCF7 ER ChIP-seq reads visualized below each figure suggest the hotspot mutations could be potentially located at tail of ER binding sites.



**Figure 20. Visualization of ENCODE data around non-coding mutations sites.**  
 The genomic regions surrounding non-coding mutations are marked by histone proteins and DNaseI hypersensitivity assay which are enriched in regulatory sites of the genome.

### GPR126 mutations

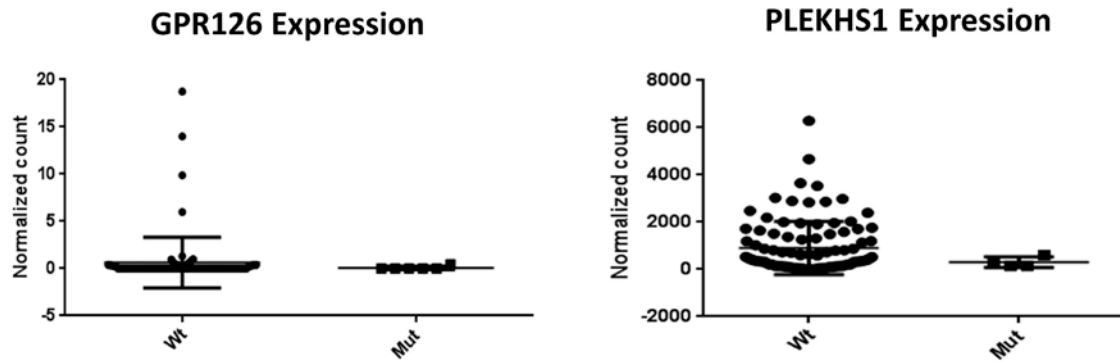
CTCTTTGTAT**GAA**CATACAAAGAG

### PLEKHS1 mutations

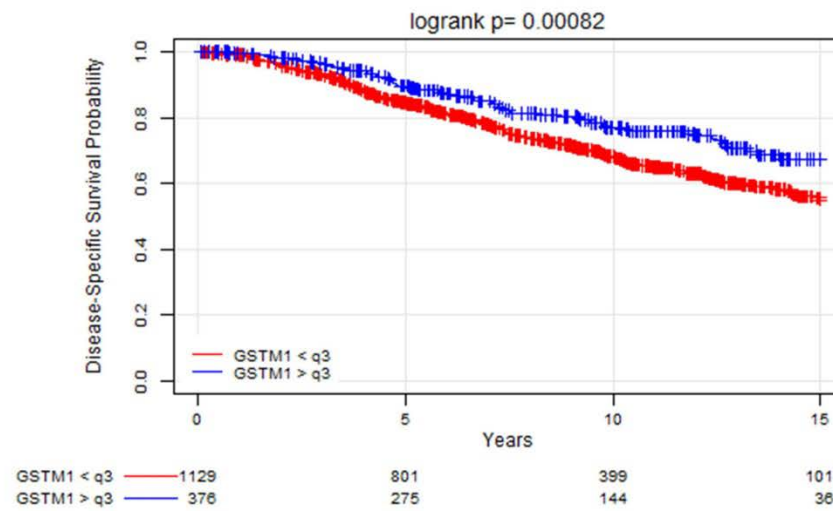
TTTTGCAATT**GAA**CAATTGCAAAA

**Figure 21. Palindromic sequence surrounding the non-coding mutations in GPR126 and PLEKHS1 genes.**

The base-pair changes are colored in red.



**Figure 22. The gene expression of GPR126 and PLEKHS1 genes in WT vs Mut carriers.**  
The gene expression was obtained for tumors with WT and Mut genotype in TCGA. No significant correlation was found in the expression of GPR126 and PLEKHS1 gene between WT and Mut carriers

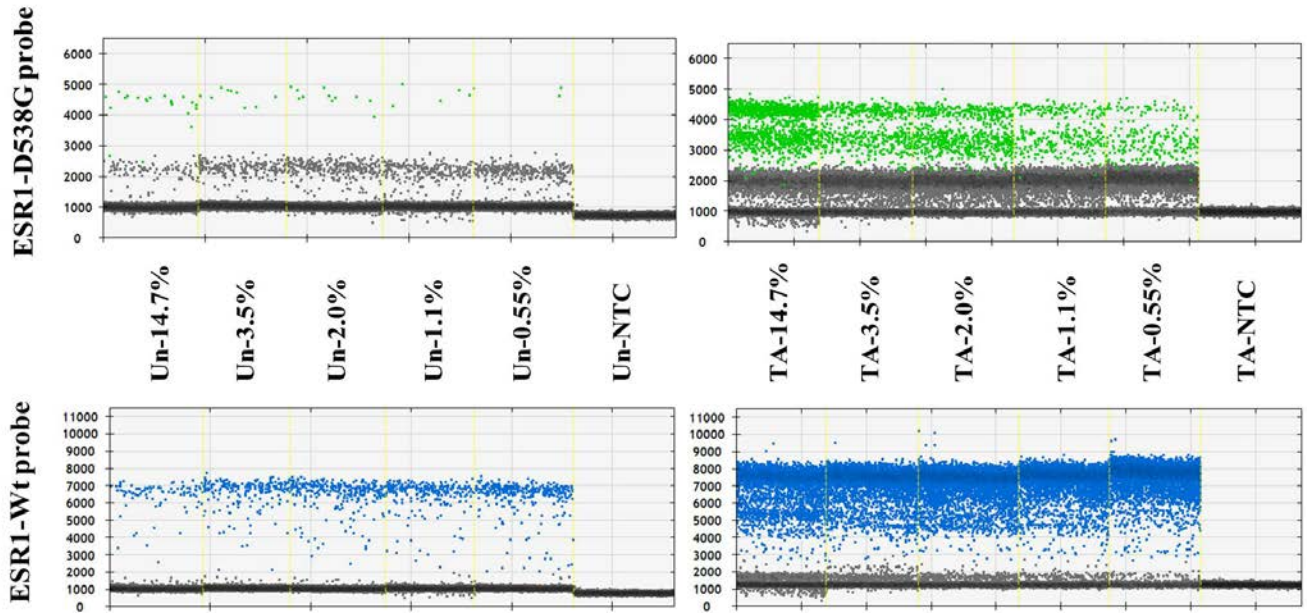


**Figure 23. Survival analysis of patients stratified by the expression of GSTM1.**

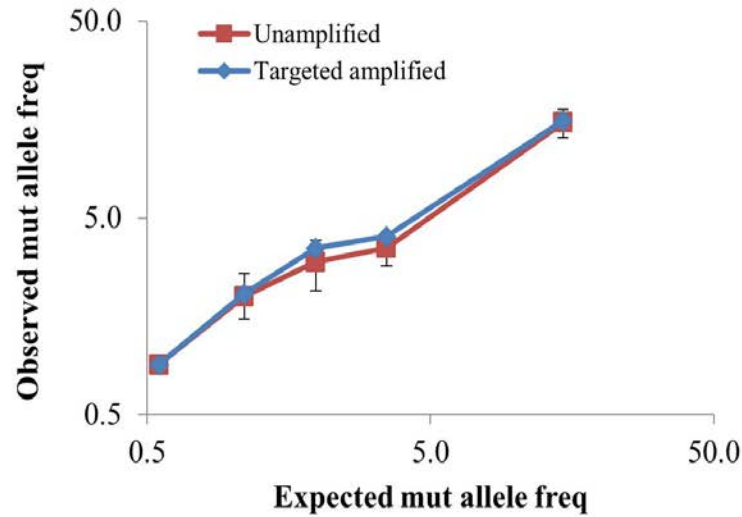
The gene expression and survival information was obtained from METABRIC data portal. The patients were classified based on the upper quartile expression of GSTM1. The patients harbouring tumors with higher expression of GSTM1 show better overall survival (in collaboration with Kevin Levine).



A



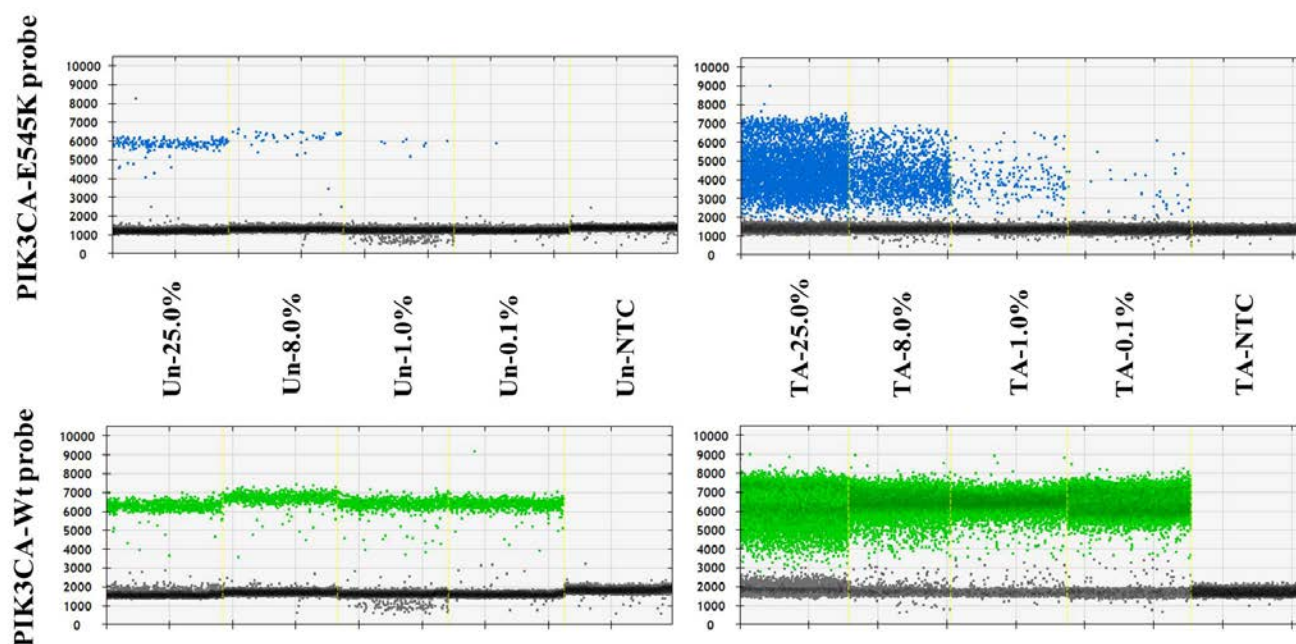
B



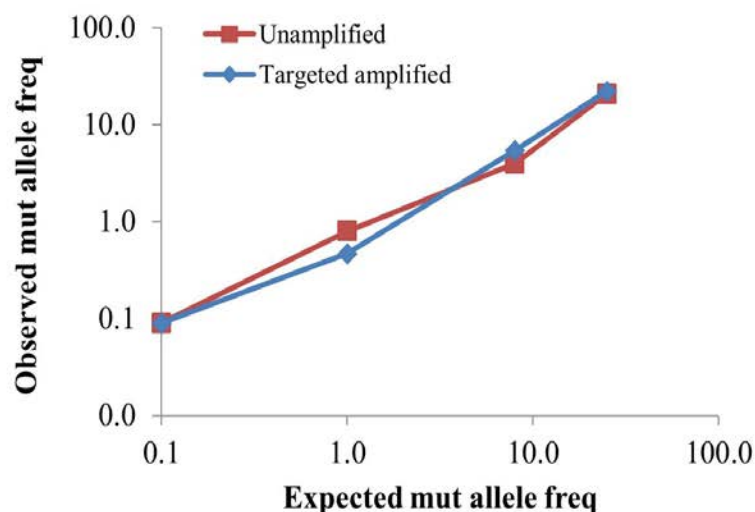
**Figure 24. Pre-amplification preserves mutant allele frequency and maintains sensitivity of *ESRI*-D538G mutation detection by ddPCR.**

2ng input DNA from admixtures of serially diluted cfDNA-16 mutant DNA and fixed amount of cfDNA-14 wildtype DNA (at resultant mutant allele frequencies of 14.7%, 3.5%, 2.0%, 1.1%, and 0.55%) were pre-amplified for 15 cycles using primers just outside of the ddPCR primers. Qiagen column-purified targeted amplified (TA) PCR products were then subjected to ddPCR analysis using 1/20<sup>th</sup> of diluted TA output along with the respective unamplified (Un) cfDNA samples (2ng input DNA). A. 1D fluorescence plots are shown for serially diluted Un and TA samples. B. Fractional mutant abundance (%) was comparable between Un and TA samples and showed linearity in serially diluted mutant cfDNA.

**A**

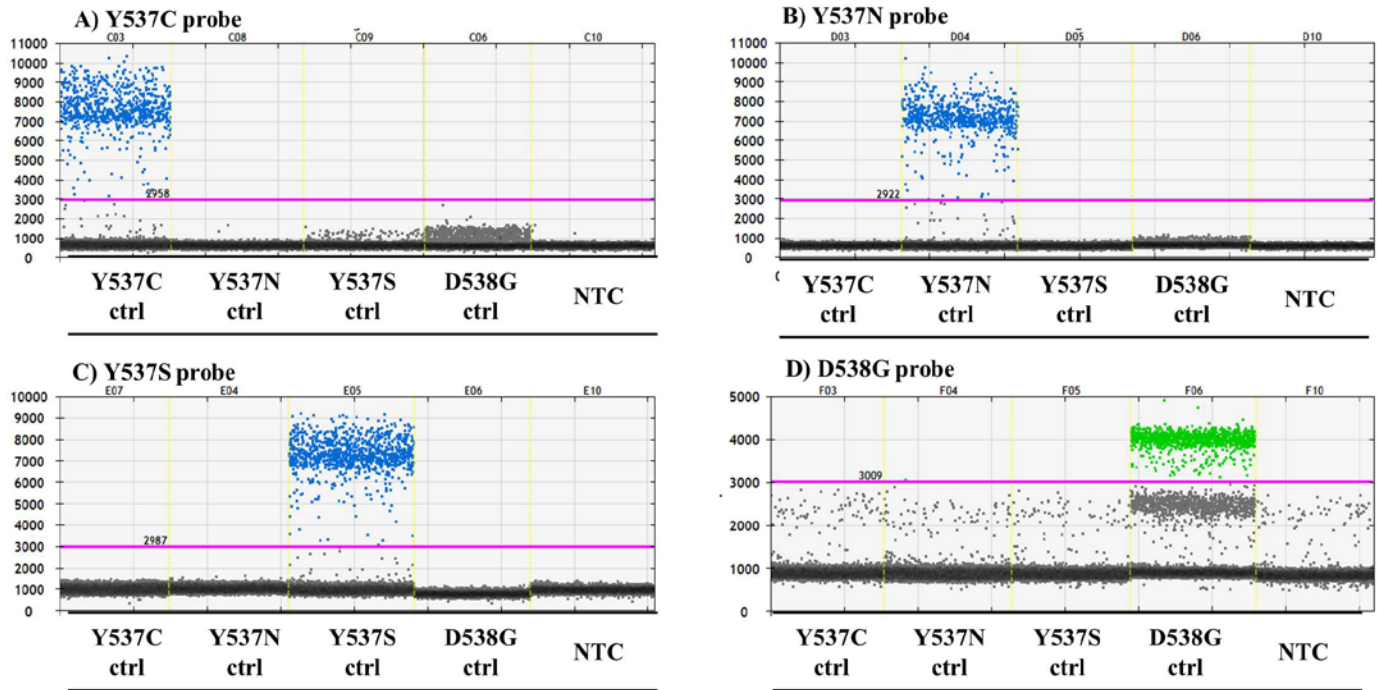


**B**



**Figure 25. Pre-amplification preserves mutant allele frequency and maintains sensitivity of PIK3CA-E545K mutation detection by ddPCR.**

1ng input DNA from admixtures of serially diluted MDA-MB-361 mutant cell line DNA and fixed amount of BT-474 wildtype cell line DNA (at resultant mutant allele frequencies of 25.0%, 8.0%, 1.0%, and 0.1%) were pre-amplified for 15 cycles using primers just outside of the ddPCR primers. Qiagen column-purified targeted amplified (TA) PCR products were then subjected to ddPCR analysis using 1/20<sup>th</sup> of diluted TA output along with the respective unamplified (Un) DNA samples (1ng input DNA). A. 1D fluorescence plots are shown for serially diluted Un and TA samples. B. Fractional mutant abundance (%) was comparable between Un and TA samples and showed linearity in serially diluted mutant DNA.

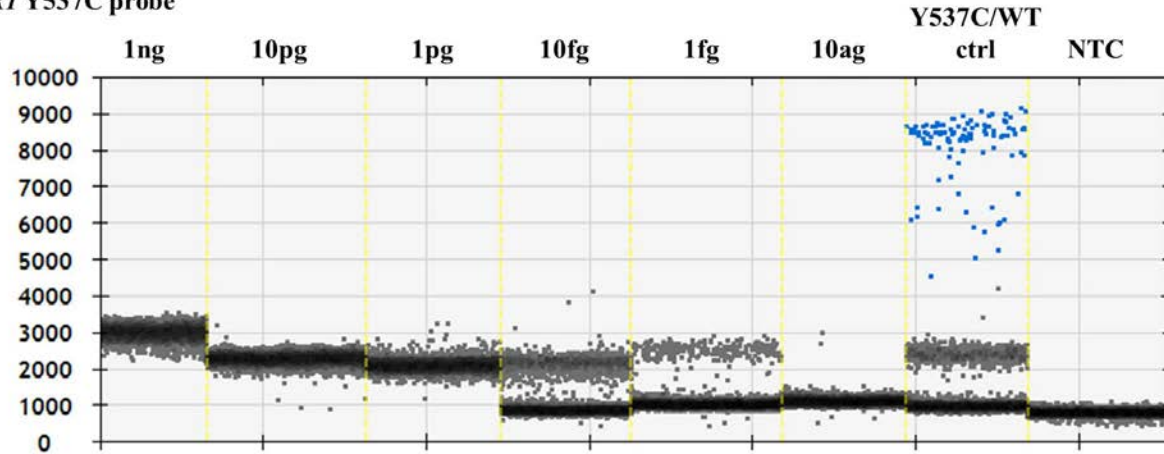


**Figure 26. *ESRI* Y537C/N/S and D538G mutation probes are specific to their corresponding mutations.**

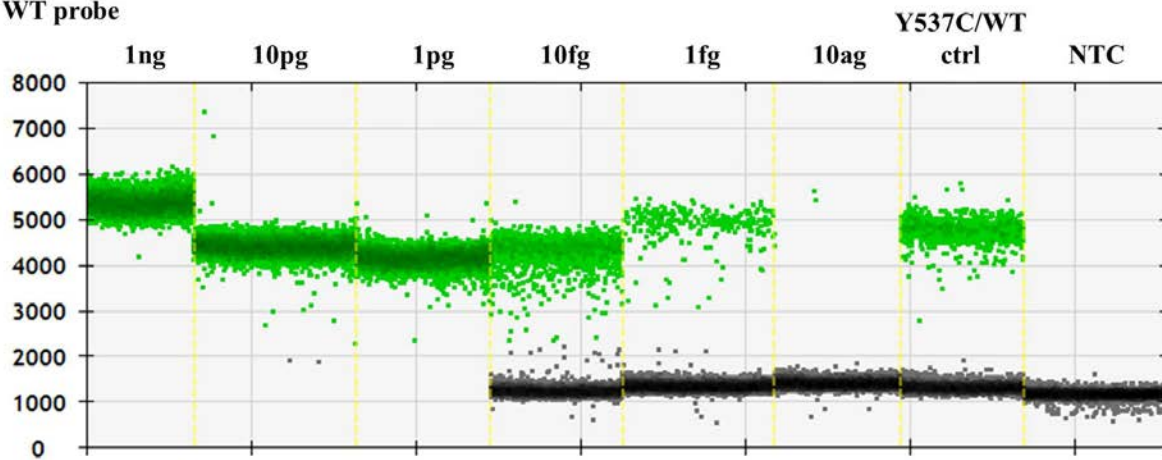
5ul of 10fM *ESRI* Y537C, Y537N, or Y537S oligo nucleotides were mixed with 30ng WT gDNA as corresponding controls. 15ng gDNA with *ESRI* D538G was used as D538G control. DNA controls with mutations as well as non-template control (NTC) were analyzed by A) Y537C probe, B) Y537N probe, C) Y537S probe, and D) D538G probe, respectively. 1D plots show mutant channel. The grey signal in D) D538G ctrl lane (fluorescence intensity ~ 2500) is for WT *ESRI* and appears only in the D538G probe as the location of WT and mutant droplet clusters are shifted inwards on the 2-D plot unlike at right angles for the remaining probes.

A)

*ESR1* Y537C probe



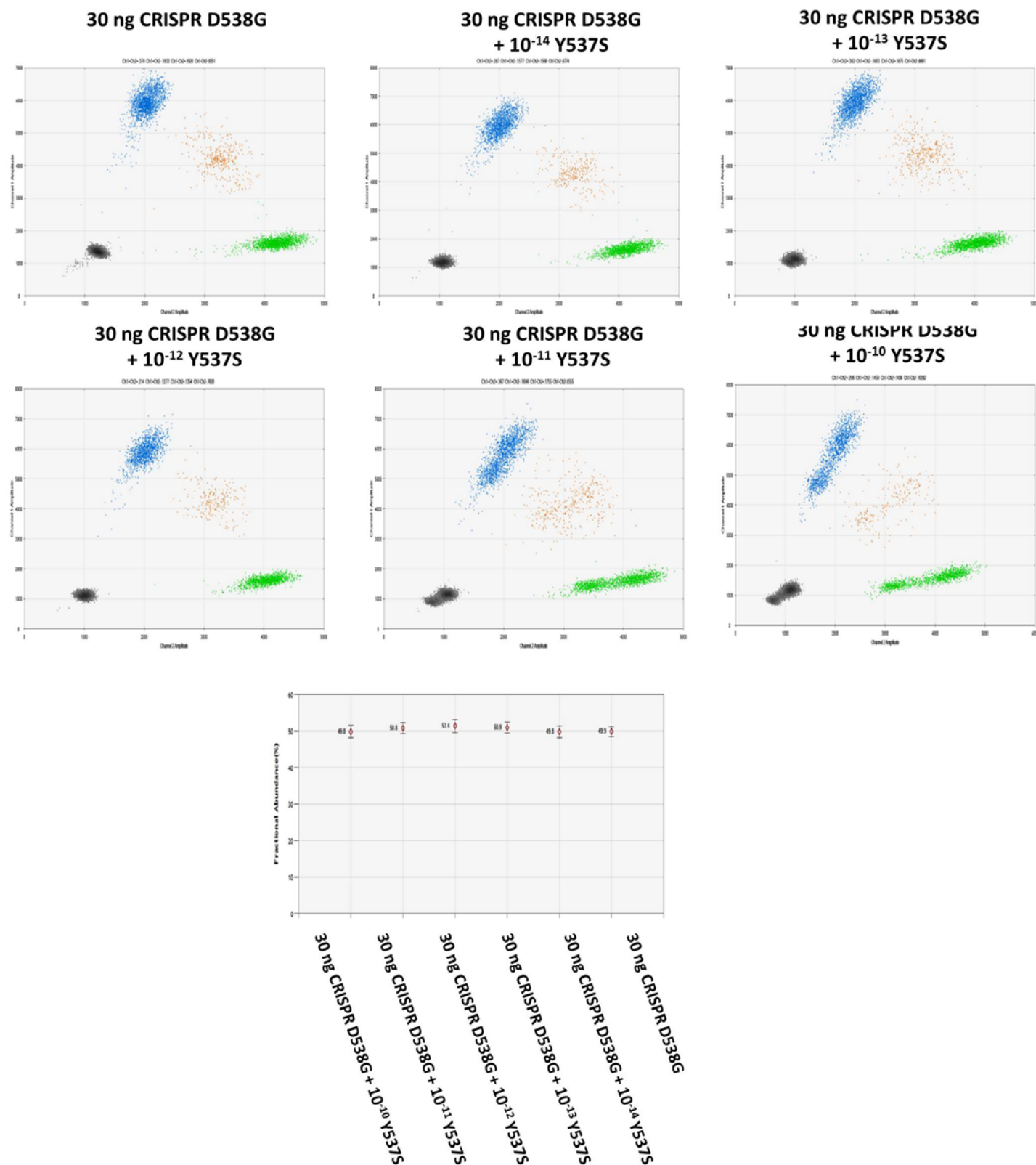
*ESR1* WT probe



**Figure 27. *ESR1* Y537C probe does not bind to wild-type allele, even at high concentrations of wild-type DNA.**

(Similar results were observed for *ESR1* D538G probe, data not shown here). 1ng, 10pg, 1pg, 10fg, 1fg, 10ag of *ESR1* WT plasmid was tested for the binding of *ESR1* Y537C mutant probe. *ESR1* Y537C oligos mixed with *ESR1* WT DNA and NTC were used as positive and negative controls, respectively. 1D plots of fluorescent signal are shown for *ESR1* Y537C (top) and WT (bottom) probes.

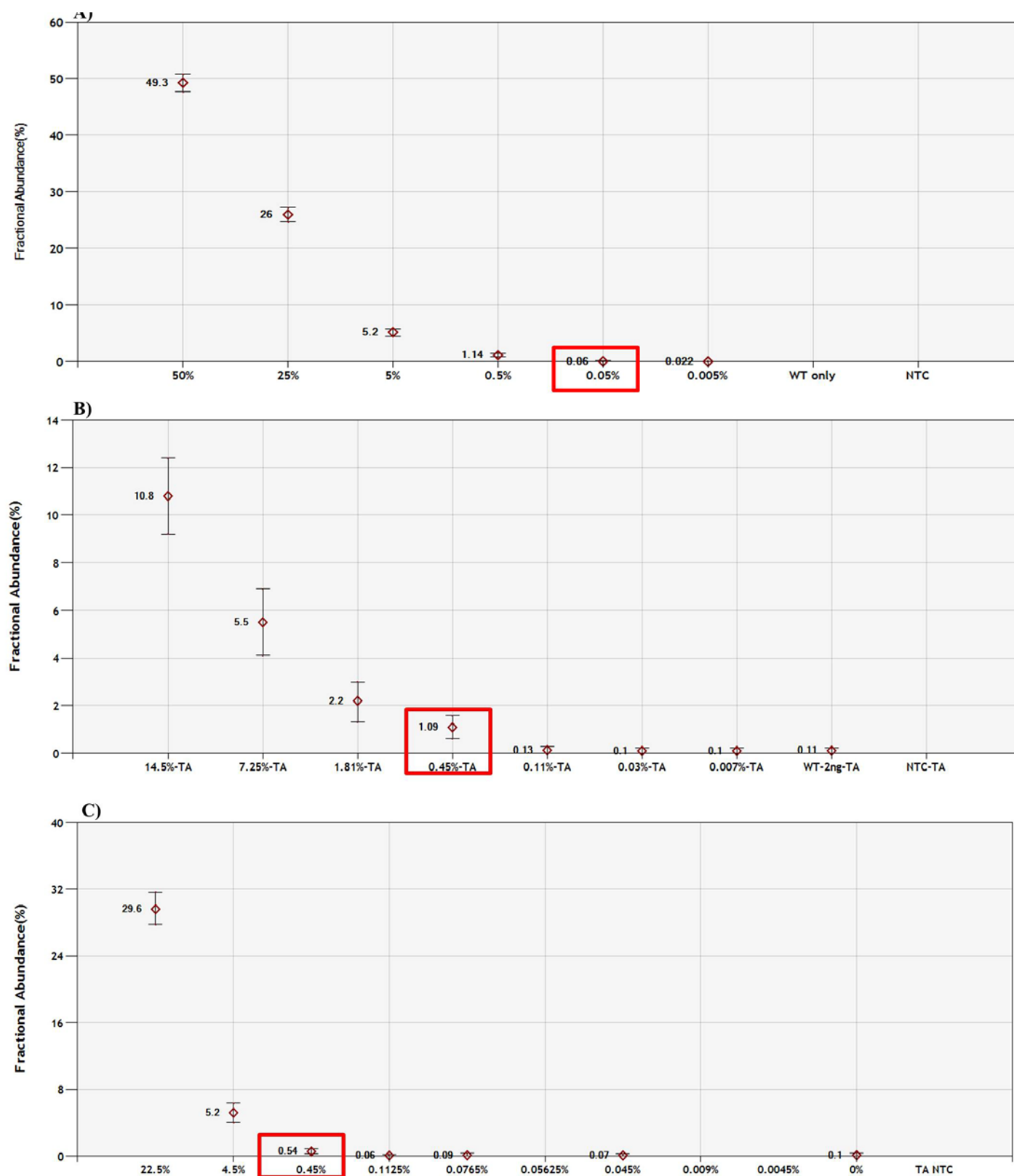
B)



**Figure 28. Lack of cross-reactivity between D538G and Y537S probes.**

Constant amount of genomic DNA (30ng) from CRISPR clone with D538G mutation was mixed with increased amount of Y537S oligo. 2D plots of fluorescent signal are shown for *ESR1* D538G and WT probes (Top). Increasing amount of Y537S did not change detected allele frequency for D538G (bottom).

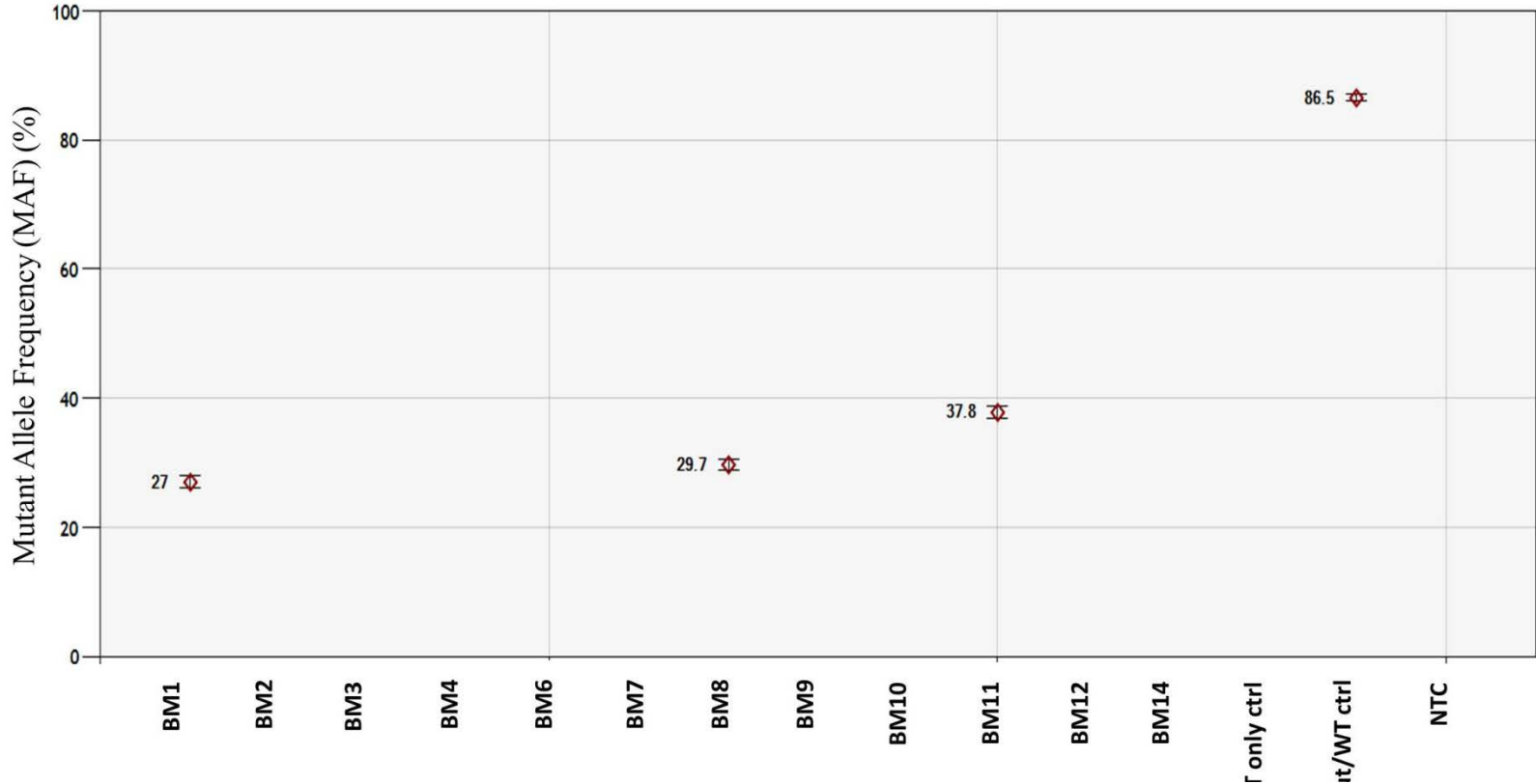




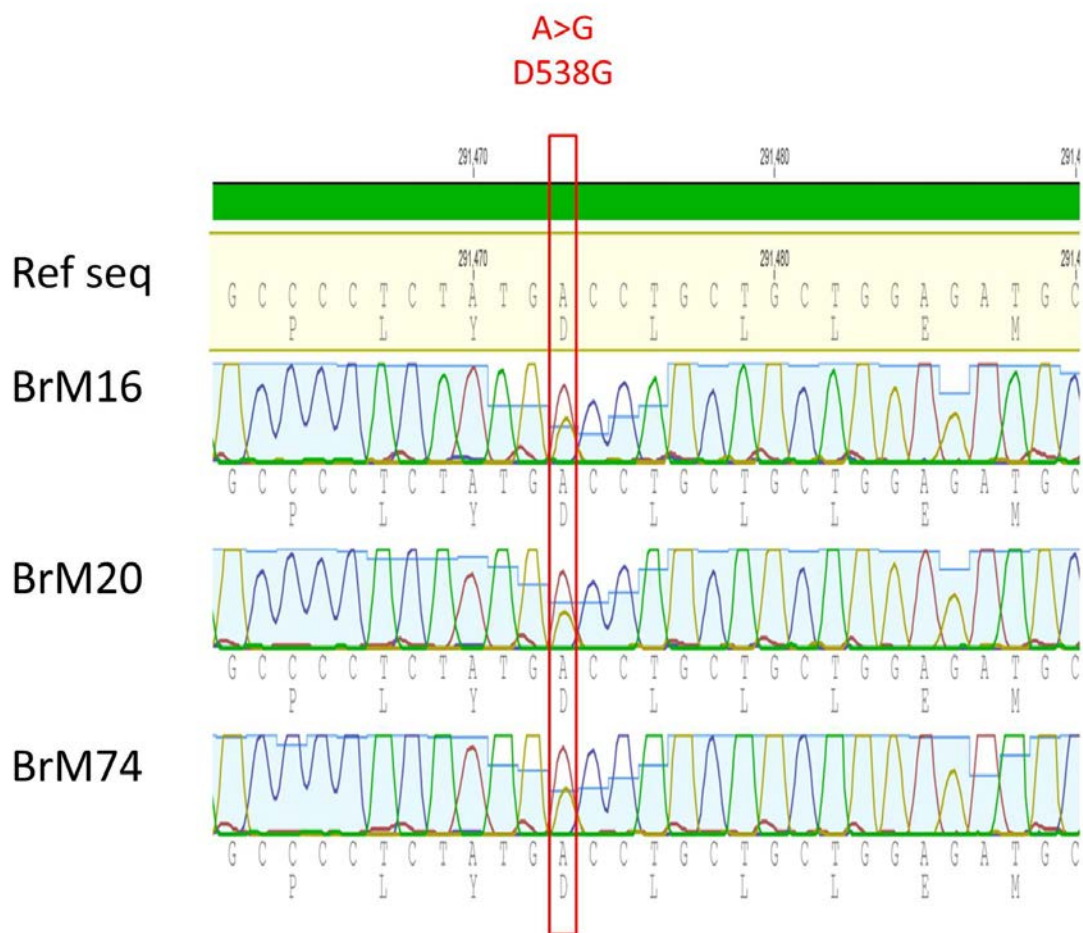
**Figure 29. LLoD determination of ddPCR**

on A) Frozen tissues, B) cfDNA, and C) FFPE tissues. *ESR1* D538G KI gDNA, *ESR1* D538G positive cfDNA, or *ESR1* D538G positive FFPE samples were spiked in *ESR1*-mutation-free frozen tissues, cfDNA, or FFPE samples respectively at different ratio. Expected allele frequency of mutations is labeled on the x axis. The LLoD were defined as the lowest mutant frequency with at least 3 droplets which was 0.05% for

frozen tissues, 0.45% for cfDNA and FFPE samples. The respective adjusted LLoD were 0.05%, 0.10%, and 0.16% after background-noise subtraction.



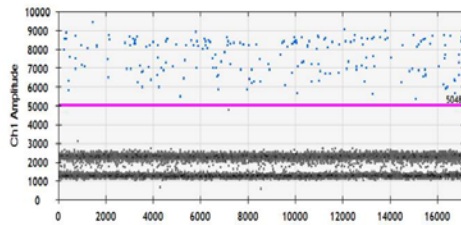
**Figure 30. Mutant allele frequency of PIK3CA H1047R mutation in 12 bone metastases.**  
The T47D cell lines DNA was used as the WT only control.



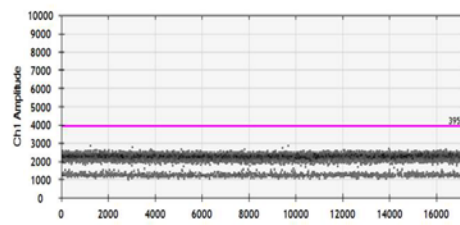
**Figure 31. The D538G ER mutation in 3 brain mets was confirmed by Sanger sequencing.**



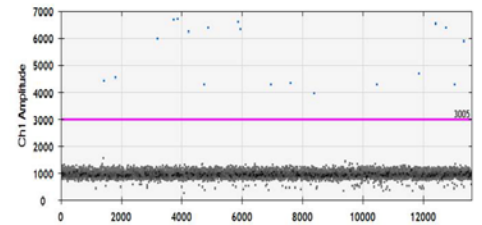
### A) *ESRI* Y537S probe



CF4 D1

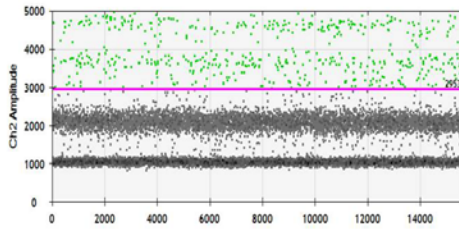


CF4 D2

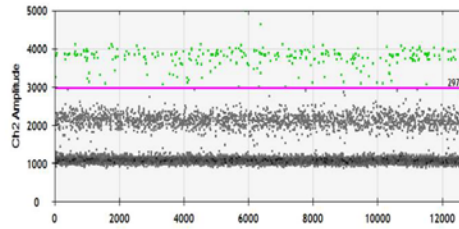


BR17

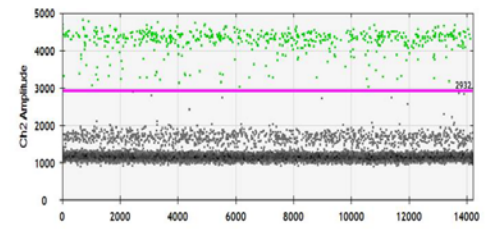
### B) *ESRI* D538G probe



CF4 D1

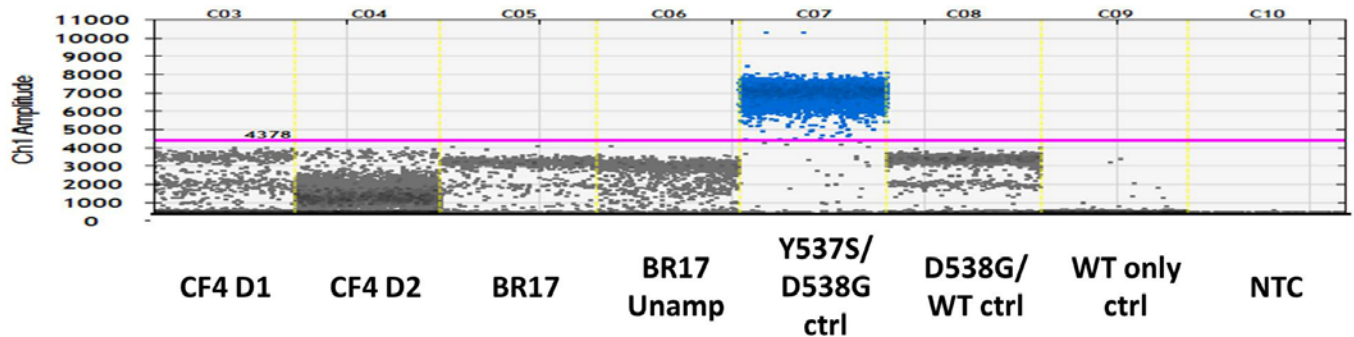


CF4 D2



BR17

### C) *ESRI* Y537S/D538G probe



### D) *ESRI* WT probe

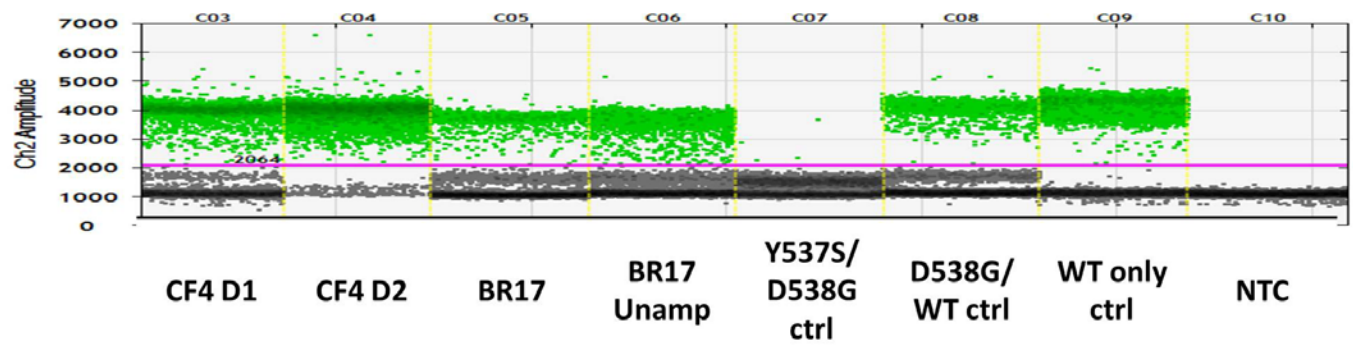
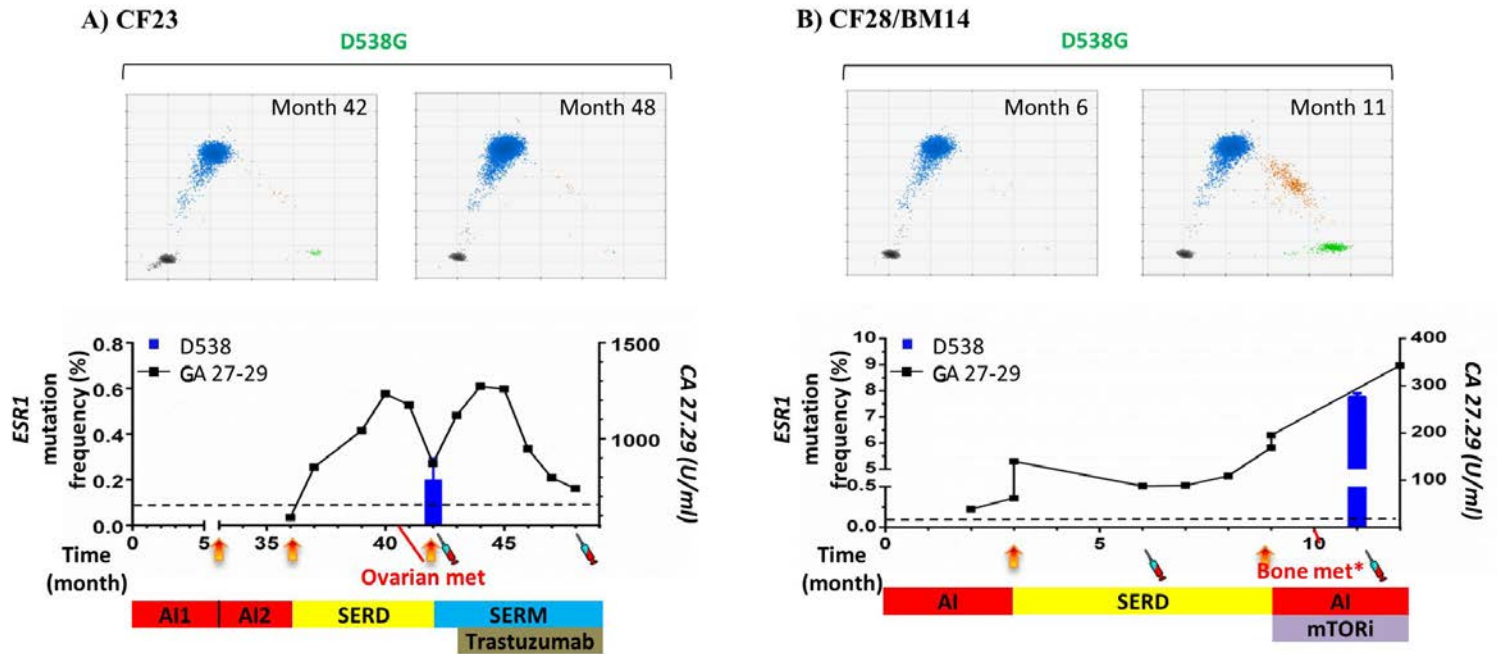


Figure 32. *ESRI* Y537S and D538G observed in the same specimens are not mutated on the same alleles.

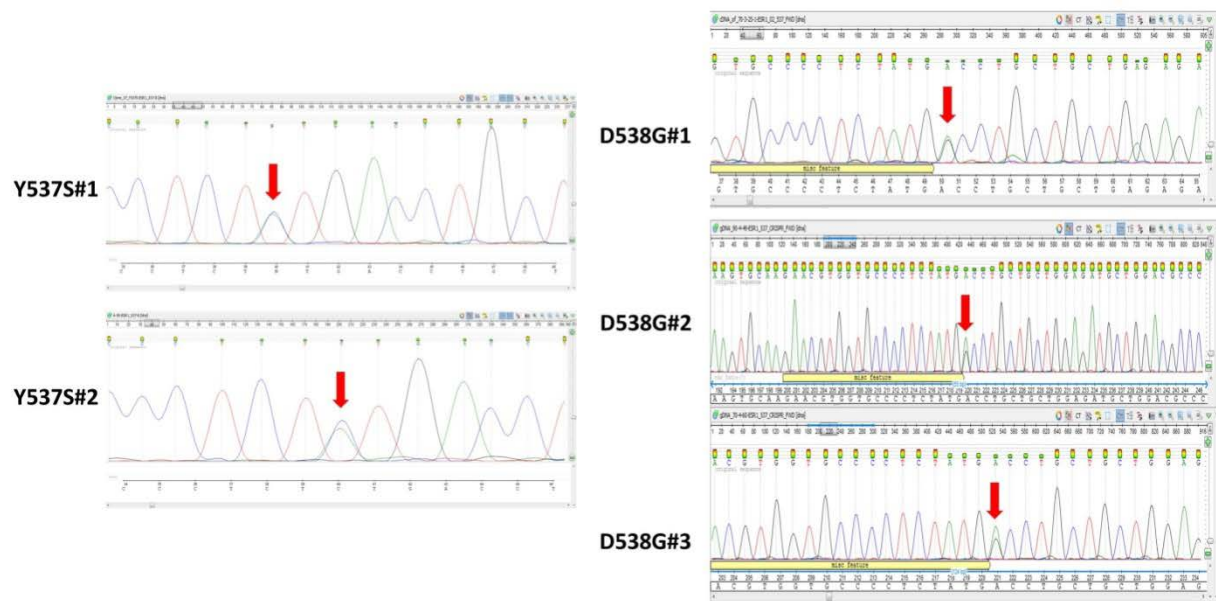
CF4 D1, CF4 D2, BR17 were tested using *ESRI* Y537S, D538G, Y537S/D538G and *ESRI* WT probes among which the latter only detects dual mutations on the same alleles at high signal amplitude.

Unamplified DNA from BR17 was also tested. Plasmid with both mutations (Y537S/D538G ctrl), gDNA with *ESR1* D538G or *ESR1* WT were used as controls.

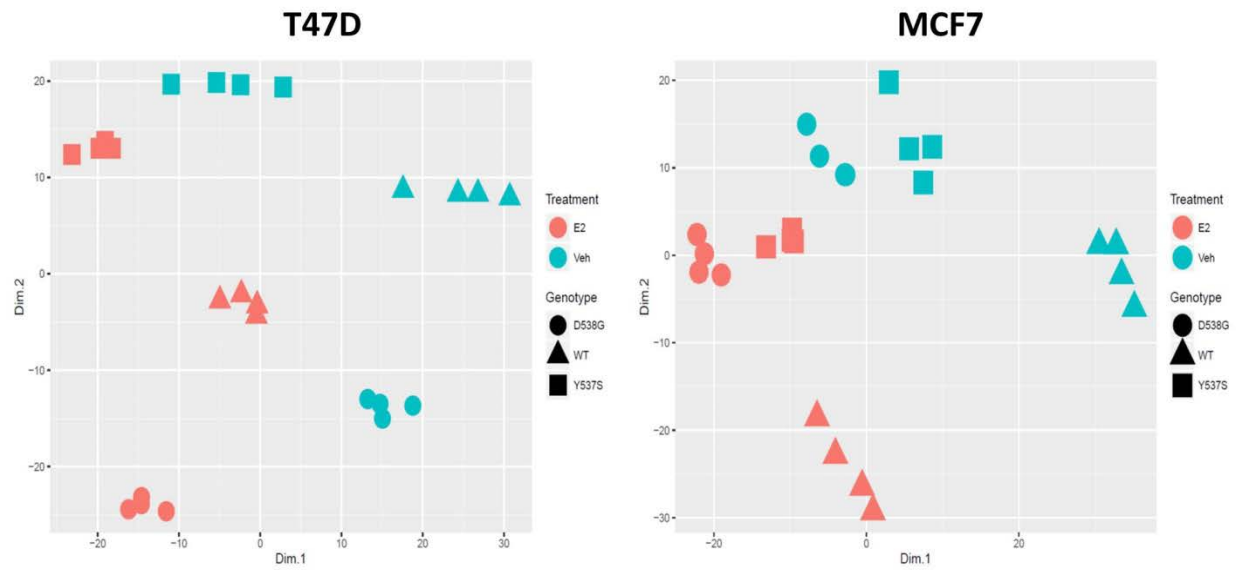


**Figure 33. Clinical timelines and allele frequency of *ESR1* mutations in multiple blood draws and matched metastatic lesions from two patients.**

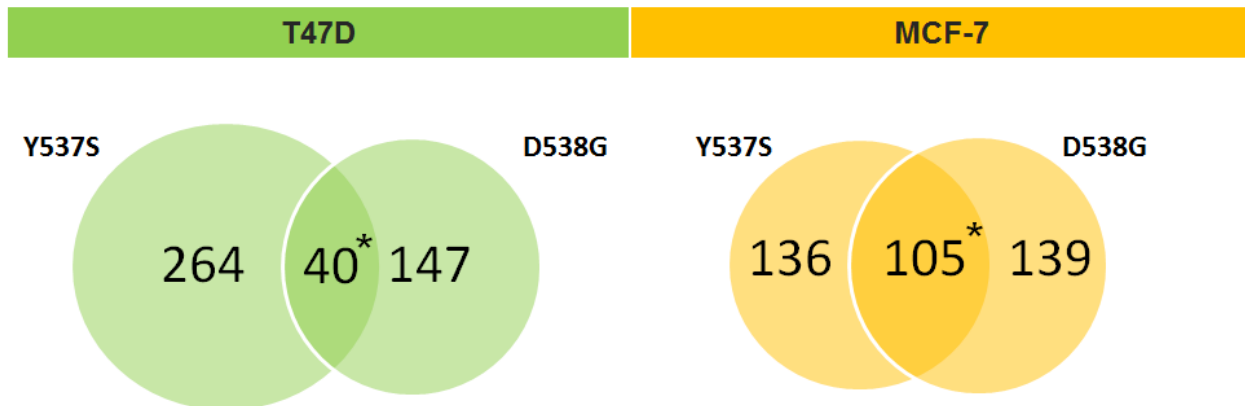
A) CF23 and B) CF28/BM14. The timeline starts with diagnosis of metastatic disease and shows treatments received, disease progression (vertical arrows), death (red cross), tumor marker assessments (CA 27-29 antigen line graph), and *ESR1* assessments (syringe and bar graphs). The mutant allele frequency of the *ESR1* mutations, measured by ddPCR, are displayed in 2D plots above the time line. The dashed line represents the LLoD (0.09%) below which the mutations were not detectable by ddPCR. Treatment abbreviations: Chemo (chemotherapy), PARPi (PARP inhibitor), LU (Leuprolide), SERM (Selective Estrogen Receptor Modulator), SERD (Selective Estrogen Receptor Degradar), AI (Aromatase Inhibitor) and mTORi (mTOR inhibitor). \*The matched metastatic lesion was positive for *ESR1* mutation. (online only).



**Figure 34. Sanger sequencing of *ESR1* mutations in T47D cells.**  
Sanger sequencing shows the insertion of Y537S (A>C) and D538G (A>G) in T47D cells.

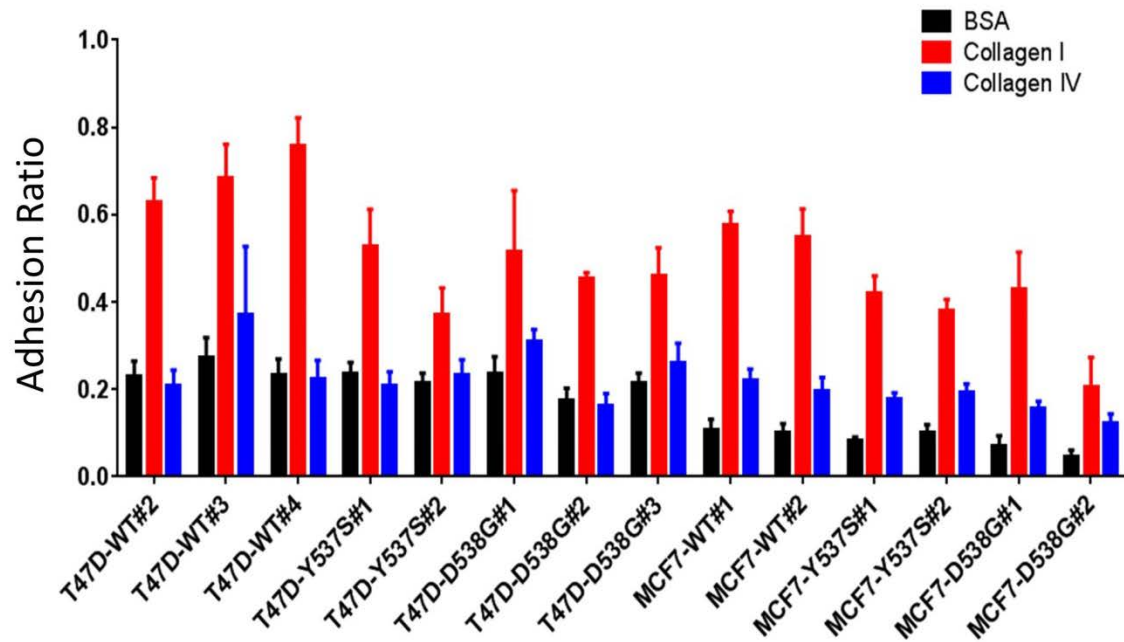


**Figure 35. PCA analysis of 1000 top variable genes between WT and mutants.**



**Figure 36. The overlap of ligand independent differentially expressed novel targets of ER.**

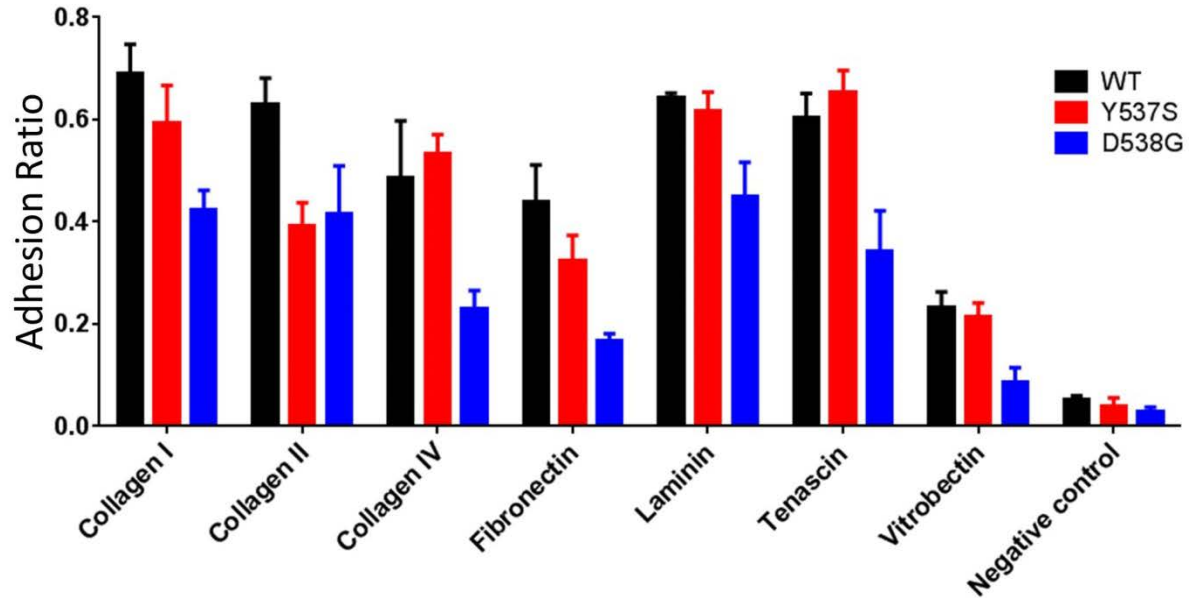
p-value<0.01



**Figure 37. Cell adhesion to Collagen I and IV in individual WT and mutant clones.**

Individual clones were grown in media+10% FBS and plated on precoated Collagen I, IV and BSA plates. The adhesion assay was performed as described in the methods.

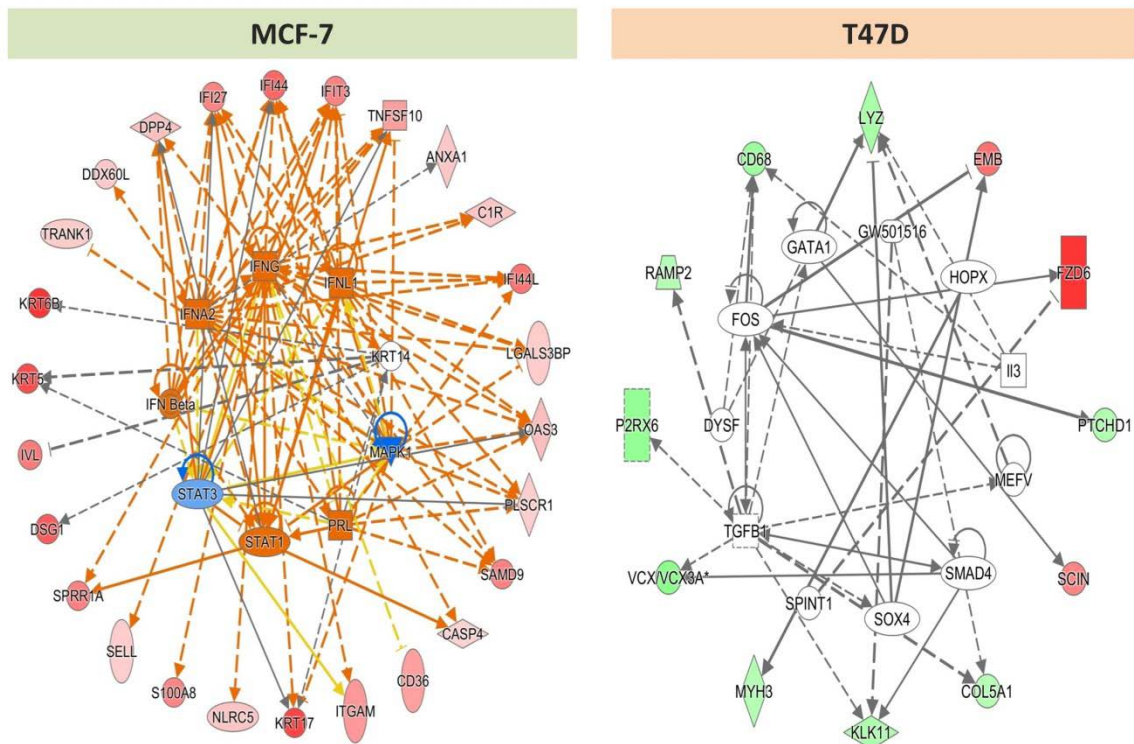
(provided by Zheqi Li)



**Figure 38. Cell adhesion different ECMs in T47D pooled WT and mutant cells.**

Individual clones were grown in media+10% FBS and pooled prior to using ECM array kit. The adhesion assay was performed as described in the methods.

(Provided by Zheqi Li)



**Figure 39. Network analysis of ligand independent novel regulated genes common between Y537S and D538G in each cell line.**

The upstream regulator analysis was performed on the genes inputted in IPA package for each cell lines independently.



## APPENDIX B: SUPPLEMENTARY TABLES

**Table 8. Primer sets used for different assays**

Assay name	Forward Sequence	Reverse Sequence
ChIP-qPCR	AGGATCCAGGCAGAGTACAG	CCCAGCCCTGTGAGCTTTA
Allele specific ChIP	GCAACACCTGGCTTCTGT	CAGCCCTGTGAGCTTTAACA
IGF1R expression	AGTTATCTCCGGTCTCTGAGG	TCTGTGGACGAACTTATTGGC
Cloning for luciferase assay	AATTGATATCCCACAGCTATGCCACCTG	AATTAAGCTTCCACAACACACCTCCCTAAT

**Table 9. The list of all ER ChIP-seq data sets in breast cancer**

Cell line/Tumor	Data set	Reads	Peaks	Ref
MCF7	1	6,442,721	37,057	(3)
	2	15,450,131	79,978	(45)
	3	20,876,237	15,677	(45)
	4	21,220,101	24,823	(45)
	5	22,924,491	30,130	(5)
	6	26,445,865	48,243	(46)
	7	26,825,998	79,731	(4)
	8	27,238,465	49,231	(4)
	9	34,587,361	66,483	(4)
T47D	1	20974865	20405	(4)
	2	71986040	8,044	(4)
	3	14490194	4,806	(3)
	4	29801890	3,098	(44)
	5	16753735	11,145	(44)
BT474	1	20830555	36,565	(4)
	2	21056403	36,337	(4)
ZR75	1	19678686	57,828	(4)
	2	62581081	34,870	(4)
	3	18981841	12,137	(44)
	4	18746193	26,820	(44)
TAMR	1	22555819	61,536	(4)
	2	29699074	46,420	(4)
MDA-MB-134	1	84820062	3,398	(93)
Good Prognosis Tumors	1	11498805	11,130	(4)
	2	6201437	13,543	(4)
	3	13907908	18,635	(4)
	4	13946259	35,788	(4)
	5	7177205	3,008	(4)
	6	15274041	6,976	(4)
	7	12566868	8,898	(4)
	8	21862228	7,167	(4)
	9	18053501	2,918	(4)
Bad Prognosis Tumors	1	14269376	14,029	(4)
	2	17447408	60,874	(4)

	Table 9 Continued			
	3	18669346	8,981	(4)
	4	15743920	36,698	(4)
	5	13886222	18,000	(4)
	6	13205910	15,181	(4)
	7	19265908	2,355	(4)
	8	18277065	5,175	(4)
	9	22858543	3,009	(4)
Metastatic samples	1	10970298	13,676	(4)
	2	17325390	79,516	(4)
	3	19851135	15,861	(4)

**Table 10. The list of regulatory SNVs in MCF7 Cell line**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr1	147476254	intergenic	NBPF25P,LOC388692	dist=99905;dist=77500	rs111816629	0	DECREASE
chr17	68148267	upstream	LINC00511	NA	rs112860966	0	DECREASE
chr16	83735945	ncRNA_intronic	LOC400548	NA	rs4783134	0	INCREASE
chr1	154453258	intronic	PMF1,PMF1-BGLAP	NA	rs2475757	9.22878E-07	INCREASE
chr8	128992864	ncRNA_intronic	PVT1	NA	chr8:128992864	2.25786E-06	INCREASE
chr17	4425801	intergenic	GGT6,SMTNL2	dist=15161;dist=8224	rs10852864	2.25786E-06	INCREASE
chr2	177210984	ncRNA_splicing	LOC102724224	NR_110599:exon1:c.255+2C>T	rs2969356	7.56077E-06	DECREASE
chr1	143643443	UTR5	PDE4DIP	NA	rs1324349	7.77221E-06	INCREASE
chr5	174111380	ncRNA_exonic	MIR4634	NA	rs7709117	8.46441E-06	DECREASE
chr12	154319	exonic	IQSEC3	NA	rs216230	1.06917E-05	INCREASE
chr16	79766256	intronic	PKD1L2	NA	rs935929	1.06917E-05	DECREASE
chr14	20387049	intergenic	RNASE1,RNASE3	dist=46173;dist=42353	rs28419520	1.09518E-05	DECREASE
chr10	94821513	intergenic	CYP26C1,CYP26A1	dist=3069;dist=1709	rs68040629	1.09518E-05	INCREASE
chr11	112083748	intergenic	LOC387810,LOC101928847	dist=152013;dist=251464	rs2055936	1.23128E-05	INCREASE
chr3	198870842	intergenic	LOC220729,KIAA0226	dist=31693;dist=11820	rs145563991	1.26161E-05	INCREASE
chr1	150165373	intergenic	THEM4,S100A10	dist=16388;dist=56637	rs2999541	1.26161E-05	DECREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr10	125987106	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	125987106	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	125987106	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr7	155348454	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr7	155348454	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr7	155348454	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE

**Table 10 continued**

chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE
chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE
chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE
chr16	317624	intronic	AXIN1	NA	rs10903014	2.53932E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr17	78165542	downstream	WDR45B	NA	chr17:78165542	3.09647E-05	DECREASE
chr17	78165542	downstream	WDR45B	NA	chr17:78165542	3.09647E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr7	43255038	intronic	HECW1	NA	rs56040296	3.28554E-05	INCREASE
chr10	121292409	upstream	RGS10	NA	rs10787978	3.38577E-05	INCREASE
chr6	157157941	intronic	ARID1B	NA	rs12208040	3.62837E-05	INCREASE
chr6	157157941	intronic	ARID1B	NA	rs12208040	3.62837E-05	INCREASE
chr6	157157941	intronic	ARID1B	NA	rs12208040	3.62837E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr13	38154404	intergenic	LINC00366,FREM2	dist=102751;dist=4769	rs2496419	4.01757E-05	INCREASE
chr13	38154404	intergenic	LINC00366,FREM2	dist=102751;dist=4769	rs2496419	4.01757E-05	INCREASE
chr13	38154404	intergenic	LINC00366,FREM2	dist=102751;dist=4769	rs2496419	4.01757E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr9	139438247	intronic	NOXA1	NA	rs11497278	4.38072E-05	INCREASE
chr20	4090809	intronic	SMOX	NA	rs13040038	4.38072E-05	DECREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE
chr7	84893979	intergenic	SEMA3D,GRM3	dist=304796;dist=1217187	rs1608484	4.7029E-05	INCREASE

**Table 10 continued**

chr7	84893979	intergenic	SEMA3D,GRM3	dist=304796;dist=1217187	rs1608484	4.7029E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr7	130961351	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr7	130961351	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr16	76642909	intergenic	CLEC3A,WWOX	dist=19407;dist=47902	rs2344922	5.17572E-05	INCREASE
chr16	76642909	intergenic	CLEC3A,WWOX	dist=19407;dist=47902	rs2344922	5.17572E-05	INCREASE
chr16	76642909	intergenic	CLEC3A,WWOX	dist=19407;dist=47902	rs2344922	5.17572E-05	INCREASE
chr1	186668364	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr12	100615617	UTR5	CHPT1	NA	chr12:100615617	5.34422E-05	DECREASE
chr12	100615617	UTR5	CHPT1	NA	chr12:100615617	5.34422E-05	DECREASE
chr17	54818764	intronic	YPEL2	NA	rs8073731	5.44256E-05	DECREASE
chr17	54818764	intronic	YPEL2	NA	rs8073731	5.44256E-05	DECREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr14	105141141	intergenic	TMEM121,MIR8071-2	dist=73557;dist=17357	rs4983455	5.69393E-05	DECREASE
chr14	105141141	intergenic	TMEM121,MIR8071-2	dist=73557;dist=17357	rs4983455	5.69393E-05	DECREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE

**Table 10 continued**

chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chr4	7379317	intronic	SORCS2	NA	rs3864203	6.09368E-05	DECREASE
chr4	7379317	intronic	SORCS2	NA	rs3864203	6.09368E-05	DECREASE
chr4	7379317	intronic	SORCS2	NA	rs3864203	6.09368E-05	DECREASE
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE
chr4	7379510	intronic	SORCS2	NA	rs3900741	6.27053E-05	INCREASE
chr4	7379510	intronic	SORCS2	NA	rs3900741	6.27053E-05	INCREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	6.2727E-05	DECREASE
chr7	84893979	intergenic	SEMA3D,GRM3	dist=304796;dist=1217187	rs1608484	6.2727E-05	INCREASE
chr2	132742681	intergenic	ANKRD30BL,GPR39	dist=10669;dist=147936	rs75955051	6.49819E-05	INCREASE

**Table 11. The list of regulatory SNVs in BT474 Cell line**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr19	18253643	upstream	JUND,MIR3188	NA	rs41523455	7.56077E-06	DECREASE
chr19	18253642	upstream	JUND,MIR3188	NA	rs41519246	7.77221E-06	DECREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr6	1033964	ncRNA_intronic	LOC285768	NA	rs7770094	1.58627E-05	DECREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr16	317624	intronic	AXIN1	NA	rs10903014	2.53932E-05	INCREASE
chr7	33839193	intergenic	BBS9,BMPER	dist=226988;dist=71855	rs6961339	2.54502E-05	INCREASE
chr7	33839193	intergenic	BBS9,BMPER	dist=226988;dist=71855	rs6961339	2.54502E-05	INCREASE
chr7	33839193	intergenic	BBS9,BMPER	dist=226988;dist=71855	rs6961339	2.54502E-05	INCREASE
chr22	43485450	intronic	PRR5,PRR5-ARHGAP8	NA	rs9614562	2.57312E-05	INCREASE
chr22	43485450	intronic	PRR5,PRR5-ARHGAP8	NA	rs9614562	2.57312E-05	INCREASE
chr22	43485450	intronic	PRR5,PRR5-ARHGAP8	NA	rs9614562	2.57312E-05	INCREASE
chr5	363825	intronic	AHRR,PDCD6	NA	chr5:363825	2.82816E-05	DECREASE
chr5	363825	intronic	AHRR,PDCD6	NA	chr5:363825	2.82816E-05	DECREASE
chr19	50041017	upstream	PVRL2	NA	rs77241309	3.77088E-05	INCREASE
chr19	50041017	upstream	PVRL2	NA	rs77241309	3.77088E-05	INCREASE
chr7	67991068	intergenic	LOC102723427,LOC100507468	dist=855956;dist=707991	chr7:67991068	4.01757E-05	INCREASE
chr7	67991068	intergenic	LOC102723427,LOC100507468	dist=855956;dist=707991	chr7:67991068	4.01757E-05	INCREASE
chr7	67991068	intergenic	LOC102723427,LOC100507468	dist=855956;dist=707991	chr7:67991068	4.01757E-05	INCREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE



**Table 11 continued**

chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	4.49834E-05	DECREASE
chrX	38744889	intergenic	MID1IP1,LINC01281	dist=194162;dist=304265	rs11797866	4.7029E-05	DECREASE
chrX	38744889	intergenic	MID1IP1,LINC01281	dist=194162;dist=304265	rs11797866	4.7029E-05	DECREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chrX	38744889	intergenic	MID1IP1,LINC01281	dist=194162;dist=304265	rs11797866	6.09368E-05	DECREASE
chr5	363825	intronic	AHRR,PDCD6	NA	chr5:363825	6.2727E-05	DECREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.55787E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE
chr14	69939953	intronic	SYNJ2BP,SYNJ2BP-COX16	NA	rs10140263	7.24677E-05	INCREASE
chr14	69939953	intronic	SYNJ2BP,SYNJ2BP-COX16	NA	rs10140263	7.24677E-05	INCREASE
chr14	69939953	intronic	SYNJ2BP,SYNJ2BP-COX16	NA	rs10140263	7.24677E-05	INCREASE
chr19	52016022	intergenic	SLC1A5,SNAR-E	dist=32340;dist=9660	rs62136763	7.25674E-05	DECREASE
chr19	52016022	intergenic	SLC1A5,SNAR-E	dist=32340;dist=9660	rs62136763	7.25674E-05	DECREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	8.0548E-05	INCREASE
chr19	52016022	intergenic	SLC1A5,SNAR-E	dist=32340;dist=9660	rs62136763	8.24616E-05	DECREASE
chr14	90039715	intergenic	LINC00642,TTC7B	dist=44713;dist=36970	chr14:90039715	8.25737E-05	DECREASE
chr14	90039715	intergenic	LINC00642,TTC7B	dist=44713;dist=36970	chr14:90039715	8.25737E-05	DECREASE
chr14	90039715	intergenic	LINC00642,TTC7B	dist=44713;dist=36970	chr14:90039715	8.25737E-05	DECREASE
chr19	50041017	upstream	PVRL2	NA	rs77241309	9.33126E-05	INCREASE

Table 11 continued

chr15	53854963	intergenic	PRTG,NEDD4	dist=32494;dist=51446	rs4774822	9.33484E-05	INCREASE
chr15	53854963	intergenic	PRTG,NEDD4	dist=32494;dist=51446	rs4774822	9.33484E-05	INCREASE
chr15	53854963	intergenic	PRTG,NEDD4	dist=32494;dist=51446	rs4774822	9.33484E-05	INCREASE
chr8	1.25E+08	intronic	FER1L6	NA	rs10089757	9.83528E-05	DECREASE
chr8	1.25E+08	intronic	FER1L6	NA	rs10089757	9.83528E-05	DECREASE
chr8	1.25E+08	intronic	FER1L6	NA	rs10089757	9.83528E-05	DECREASE
chr9	1.28E+08	downstream	MVB12B	NA	rs3739564	0.00010804	INCREASE
chr9	1.28E+08	downstream	MVB12B	NA	rs3739564	0.00010804	INCREASE
chr10	97657673	ncRNA_intronic	ENTPD1-AS1	NA	rs7906654	0.000108627	INCREASE
chr10	97657673	ncRNA_intronic	ENTPD1-AS1	NA	rs7906654	0.000108627	INCREASE
chr10	97657673	ncRNA_intronic	ENTPD1-AS1	NA	rs7906654	0.000108627	INCREASE
chr10	1.15E+08	intergenic	TCF7L2,HABP2	dist=313308;dist=69846	rs2036551	0.000108702	DECREASE
chr10	1.15E+08	intergenic	TCF7L2,HABP2	dist=313308;dist=69846	rs2036551	0.000108702	DECREASE
chr11	10366316	ncRNA_intronic	CAND1.11	NA	rs77384703	0.000108702	DECREASE
chr11	10366316	ncRNA_intronic	CAND1.11	NA	rs77384703	0.000108702	DECREASE
chr5	67728553	intergenic	PIK3R1,SLC30A5	dist=95148;dist=696979	chr5:67728553	0.00011304	INCREASE
chr5	67728553	intergenic	PIK3R1,SLC30A5	dist=95148;dist=696979	chr5:67728553	0.00011304	INCREASE
chr20	25106862	intergenic	LOC284798,LOC101926889	dist=29436;dist=6446	rs118015381	0.00011304	INCREASE
chr20	25106862	intergenic	LOC284798,LOC101926889	dist=29436;dist=6446	rs118015381	0.00011304	INCREASE
chr2	11712184	intergenic	GREB1,NTSR2	dist=11821;dist=3571	rs6432223	0.00011304	INCREASE
chr2	11712184	intergenic	GREB1,NTSR2	dist=11821;dist=3571	rs6432223	0.00011304	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	0.0001149	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	0.0001149	INCREASE
chr1	1.76E+08	intergenic	BRINP2,LOC101928778	dist=166441;dist=251660	rs2207028	0.0001149	INCREASE
chr1	1.76E+08	intergenic	BRINP2,LOC101928778	dist=166441;dist=251660	rs2207028	0.0001149	INCREASE
chr1	1.76E+08	intergenic	BRINP2,LOC101928778	dist=166441;dist=251660	rs2207028	0.0001149	INCREASE
chr5	1.33E+08	intergenic	FSTL4,C5orf15	dist=272348;dist=70627	rs10736848	0.000116617	INCREASE
chr5	1.33E+08	intergenic	FSTL4,C5orf15	dist=272348;dist=70627	rs10736848	0.000116617	INCREASE
chr5	1.33E+08	intergenic	FSTL4,C5orf15	dist=272348;dist=70627	rs10736848	0.000116617	INCREASE

**Table 11 continued**

chr7	16801076	intronic	AGR2	NA	rs4719480	0.000117929	DECREASE
chr7	16801076	intronic	AGR2	NA	rs4719480	0.000117929	DECREASE
chr7	16801076	intronic	AGR2	NA	rs4719480	0.000117929	DECREASE
chr21	45203197	intronic	FAM207A	NA	rs2877001	0.000122819	INCREASE
chr21	45203197	intronic	FAM207A	NA	rs2877001	0.000122819	INCREASE
chr21	45203197	intronic	FAM207A	NA	rs2877001	0.000122819	INCREASE
chr2	48830624	intronic	LHCGR,STON1-GTF2A1L	NA	rs4555391	0.000122819	INCREASE
chr2	48830624	intronic	LHCGR,STON1-GTF2A1L	NA	rs4555391	0.000122819	INCREASE
chr2	48830624	intronic	LHCGR,STON1-GTF2A1L	NA	rs4555391	0.000122819	INCREASE
chr11	1.01E+08	intergenic	KIAA1377,C11orf70	dist=10313;dist=36060	rs10791548	0.000123861	DECREASE
chr11	1.01E+08	intergenic	KIAA1377,C11orf70	dist=10313;dist=36060	rs10791548	0.000123861	DECREASE
chr17	78426666	intronic	TBCD	NA	rs8078446	0.000124464	DECREASE
chr17	78426666	intronic	TBCD	NA	rs8078446	0.000124464	DECREASE
chr17	78426666	intronic	TBCD	NA	rs8078446	0.000124464	DECREASE
chr1	1.46E+08	intronic	NBPF10,NBPF8	NA	chr1:146192389	0.000130941	DECREASE
chr1	1.46E+08	intronic	NBPF10,NBPF8	NA	chr1:146192389	0.000130941	DECREASE

**Table 12. The list of regulatory SNVs in MDA-MB-134 Cell line**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr15	67907815	intergenic	LOC145837,LINC00593	dist=256982;dist=6812	rs305024	1.26161E-05	INCREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.37513E-05	INCREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.37513E-05	INCREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.72702E-05	INCREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr3	65769158	intronic	MAGI1	NA	rs2017783	3.2491E-05	INCREASE
chr3	65769158	intronic	MAGI1	NA	rs2017783	3.2491E-05	INCREASE
chr3	65769158	intronic	MAGI1	NA	rs2017783	3.2491E-05	INCREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr21	10168351	intergenic	BAGE,NONE	dist=47543;dist=NONE	rs143333505	7.12551E-05	INCREASE
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	8.15262E-05	DECREASE
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	8.15262E-05	DECREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chrY	57411338	intergenic	NONE,SPRY3	dist=NONE;dist=98507	rs75814907	8.43299E-05	DECREASE

Table 12 continued

chrY	57411338	intergenic	NONE,SPRY3	dist=NONE;dist=98507	rs75814907	8.43299E-05	DECREASE
chrY	57411338	intergenic	NONE,SPRY3	dist=NONE;dist=98507	rs75814907	9.26148E-05	DECREASE
chr10	41961161	intergenic	NONE,LOC441666	dist=NONE;dist=186159	rs61847195	9.83528E-05	INCREASE
chr10	41961161	intergenic	NONE,LOC441666	dist=NONE;dist=186159	rs61847195	9.83528E-05	INCREASE
chr17	41334421	intronic	MAPT	NA	rs1560311	0.000103514	DECREASE
chr17	41334421	intronic	MAPT	NA	rs1560311	0.000103514	DECREASE
chr1	5655337	intergenic	MIR4417,MIR4689	dist=108547;dist=189982	rs75831410	0.000103834	INCREASE
chr1	5655337	intergenic	MIR4417,MIR4689	dist=108547;dist=189982	rs75831410	0.000103834	INCREASE
chr1	5655337	intergenic	MIR4417,MIR4689	dist=108547;dist=189982	rs75831410	0.000103834	INCREASE
chr7_random	191044	intergenic	NONE,LOC389831	dist=NONE;dist=14698	chr7_random:191044	0.000108627	INCREASE
chr7_random	191044	intergenic	NONE,LOC389831	dist=NONE;dist=14698	chr7_random:191044	0.000108627	INCREASE
chr9_random	613232	intergenic	NONE,RNF208	dist=NONE;dist=513406	chr9_random:613232	0.000108627	DECREASE
chr9_random	613232	intergenic	NONE,RNF208	dist=NONE;dist=513406	chr9_random:613232	0.000108627	DECREASE
chr4	4007600	ncRNA_intronic	FAM86EP	NA	rs3892453	0.000110161	DECREASE
chr4	4007600	ncRNA_intronic	FAM86EP	NA	rs3892453	0.000110161	DECREASE
chr4	4007600	ncRNA_intronic	FAM86EP	NA	rs3892453	0.000110161	DECREASE
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	0.000110161	DECREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.00011304	INCREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.00011304	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.000119481	INCREASE
chr6	1.58E+08	intergenic	ARID1B,ZDHHC14	dist=80543;dist=68397	chr6:157654148	0.000122819	INCREASE
chr6	1.58E+08	intergenic	ARID1B,ZDHHC14	dist=80543;dist=68397	chr6:157654148	0.000122819	INCREASE
chr6	1.58E+08	intergenic	ARID1B,ZDHHC14	dist=80543;dist=68397	chr6:157654148	0.000122819	INCREASE
chr1	1.44E+08	intronic	NBPF12,NBPF20,NOTCH2NL	NA	rs2794088	0.000123994	INCREASE
chr1	1.44E+08	intronic	NBPF12,NBPF20,NOTCH2NL	NA	rs2794088	0.000123994	INCREASE
chr1	1.44E+08	intronic	NBPF12,NBPF20,NOTCH2NL	NA	rs2794088	0.000123994	INCREASE

Table 12 continued

chr1	16779915	exonic	NBPF1	NA	rs78540200	0.000133177	DECREASE
chrX	1.52E+08	intronic	ZNF185	NA	rs3761504	0.000138172	INCREASE
chrX	1.52E+08	intronic	ZNF185	NA	rs3761504	0.000138172	INCREASE
chrY	10594845	intergenic	TTY23,NONE	dist=235274;dist=NONE	rs77846686	0.000138172	DECREASE
chrY	10594845	intergenic	TTY23,NONE	dist=235274;dist=NONE	rs77846686	0.000138172	DECREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000142131	INCREASE
chr20	26149877	intergenic	MIR663AHG,NONE	dist=12008;dist=NONE	rs62198169	0.000142721	DECREASE
chr20	26149877	intergenic	MIR663AHG,NONE	dist=12008;dist=NONE	rs62198169	0.000142721	DECREASE
chr20	26149877	intergenic	MIR663AHG,NONE	dist=12008;dist=NONE	rs62198169	0.000142721	DECREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr7_random	176814	intergenic	NONE,LOC389831	dist=NONE;dist=28928	chr7_random:176814	0.000146461	DECREASE
chr7_random	176814	intergenic	NONE,LOC389831	dist=NONE;dist=28928	chr7_random:176814	0.000146461	DECREASE
chr7_random	176814	intergenic	NONE,LOC389831	dist=NONE;dist=28928	chr7_random:176814	0.000146461	DECREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000146461	INCREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000146461	INCREASE
chr17	41334421	intronic	MAPT	NA	rs1560311	0.000159308	DECREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000162941	DECREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000162941	DECREASE
chr12	56403622	downstream	AGAP2	NA	rs238516	0.000165147	DECREASE
chr12	56403622	downstream	AGAP2	NA	rs238516	0.000165147	DECREASE
chr17_random	95928	intergenic	NONE,MGC70870	dist=NONE;dist=20695	chr17_random:95928	0.000169076	DECREASE
chr7_random	180317	intergenic	NONE,LOC389831	dist=NONE;dist=25425	chr7_random:180317	0.000178475	DECREASE
chr7_random	180317	intergenic	NONE,LOC389831	dist=NONE;dist=25425	chr7_random:180317	0.000178475	DECREASE
chr22	26751300	ncRNA_intronic	MIR548AM	NA	rs1548235	0.000178475	INCREASE
chr22	26751300	ncRNA_intronic	MIR548AM	NA	rs1548235	0.000178475	INCREASE
chr22	26751300	ncRNA_intronic	MIR548AM	NA	rs1548235	0.000178475	INCREASE
chr1	5654098	intergenic	MIR4417,MIR4689	dist=107308;dist=191221	rs72863341	0.000178475	DECREASE

**Table 12 continued**

chr1	5654098	intergenic	MIR4417,MIR4689	dist=107308;dist=191221	rs72863341	0.000178475	DECREASE
chr1	5654098	intergenic	MIR4417,MIR4689	dist=107308;dist=191221	rs72863341	0.000178475	DECREASE
chr2	91171920	intergenic	NONE,LOC654342	dist=NONE;dist=16516	rs2531397	0.000188278	INCREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000188278	DECREASE
chr22	49429472	intergenic	ARSA,SHANK3	dist=16005;dist=30464	rs62241523	0.000188278	INCREASE
chr22	49429472	intergenic	ARSA,SHANK3	dist=16005;dist=30464	rs62241523	0.000188278	INCREASE
chr22	49429472	intergenic	ARSA,SHANK3	dist=16005;dist=30464	rs62241523	0.000188278	INCREASE
chr2	57173824	intergenic	RNU6-35P,VRK2	dist=494952;dist=814466	rs71416469	0.000191251	DECREASE
chr2	57173824	intergenic	RNU6-35P,VRK2	dist=494952;dist=814466	rs71416469	0.000191251	DECREASE
chr2	57173824	intergenic	RNU6-35P,VRK2	dist=494952;dist=814466	rs71416469	0.000191251	DECREASE
chr7_random	174962	intergenic	NONE,LOC389831	dist=NONE;dist=30780	chr7_random:174962	0.000196412	DECREASE
chr7_random	174962	intergenic	NONE,LOC389831	dist=NONE;dist=30780	chr7_random:174962	0.000196412	DECREASE
chr8	11625848	intronic	GATA4	NA	rs10096189	0.000196412	DECREASE
chr8	11625848	intronic	GATA4	NA	rs10096189	0.000196412	DECREASE
chr12	1.21E+08	intronic	TMEM120B	NA	rs4760112	0.000196412	INCREASE
chr12	1.21E+08	intronic	TMEM120B	NA	rs4760112	0.000196412	INCREASE
chr20	28206811	intergenic	NONE,FRG1B	dist=NONE;dist=18729	rs62198925	0.000200867	DECREASE



**Table 13. The list of regulatory SNVs in T47D Cell line**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	3.46458E-05	INCREASE
chr14	94865979	ncRNA_intronic	LOC101929080	NA	rs111943653	7.58002E-05	INCREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	0.000103704	INCREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	0.000103704	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr1	17514023	intronic	PADI4	NA	rs2501808	0.000117929	INCREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000142131	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000146461	INCREASE
chr4	7098724	ncRNA_exonic	LOC100129931	NA	rs10012093	0.000146461	INCREASE
chr14	94865979	ncRNA_intronic	LOC101929080	NA	rs111943653	0.0001532	INCREASE
chr14	94865979	ncRNA_intronic	LOC101929080	NA	rs111943653	0.0001532	INCREASE
chr2	45833232	intronic	PRKCE	NA	rs73926056	0.000185539	DECREASE
chr9	76905965	intronic	OSTF1	NA	rs11144228	0.000261882	INCREASE
chr9	76905965	intronic	OSTF1	NA	rs11144228	0.000261882	INCREASE
chr6	6704104	intergenic	LY86,RREB1	dist=103889;dist=348725	rs6938081	0.000320937	DECREASE
chr15	91090470	intergenic	FAM174B,ASB9P1	dist=90435;dist=49248	rs12440115	0.000330423	DECREASE
chr15	91090470	intergenic	FAM174B,ASB9P1	dist=90435;dist=49248	rs12440115	0.000330423	DECREASE
chr15	91090470	intergenic	FAM174B,ASB9P1	dist=90435;dist=49248	rs12440115	0.000330423	DECREASE
chr5	1.32E+08	intronic	SLC22A5	NA	rs274567	0.000349852	DECREASE
chr14	93867312	intergenic	SERPINA6,SERPINA2	dist=7871;dist=32416	rs1884551	0.000411044	INCREASE



**Table 13 continued**

chr17	73929111	intronic	PGS1	NA	rs2138127	0.000609653	DECREASE
chr17	73929111	intronic	PGS1	NA	rs2138127	0.000609653	DECREASE
chr2	45833232	intronic	PRKCE	NA	rs73926056	0.000763915	DECREASE
chr2	45833232	intronic	PRKCE	NA	rs73926056	0.000763915	DECREASE
chr9	76905965	intronic	OSTF1	NA	rs11144228	0.000882191	INCREASE
chr10	82717289	intergenic	SH2D4B,NRG3	dist=320993;dist=907761	rs2881074	0.001124783	INCREASE
chr14	94865979	ncRNA_intronic	LOC101929080	NA	rs111943653	0.001188517	DECREASE
chr22	27549122	intergenic	XBP1,ZNRF3	dist=22562;dist=60633	rs6519764	0.001188517	DECREASE
chr7	34070750	intronic	BMPER	NA	rs1362456	0.001196024	DECREASE
chr15	31147775	intronic	FMN1	NA	rs11639303	0.001200055	DECREASE
chr12	51650094	intergenic	KRT18,EIF4B	dist=17142;dist=36235	rs7306508	0.001291262	INCREASE
chr12	51650094	intergenic	KRT18,EIF4B	dist=17142;dist=36235	rs7306508	0.001291262	INCREASE
chr12	51650094	intergenic	KRT18,EIF4B	dist=17142;dist=36235	rs7306508	0.001291262	INCREASE
chr16	82538204	intergenic	MLYCD,OSGIN1	dist=30916;dist=6124	rs4782862	0.001414265	INCREASE
chr11	69039535	intergenic	LOC101928292,CCND1	dist=22043;dist=125519	rs2015489	0.001620628	DECREASE
chr16	87123750	intronic	ZFPM1	NA	rs12929780	0.001745782	INCREASE
chr17	72796567	intronic	41891	NA	rs1702937	0.001815744	DECREASE
chr15	91090470	intergenic	FAM174B,ASB9P1	dist=90435;dist=49248	rs12440115	0.001906126	INCREASE
chr2	65380947	intergenic	ACTR2,SPRED2	dist=29056;dist=10542	rs6546142	0.001906126	INCREASE
chr7	36091291	intergenic	SEPT7,LOC101928618	dist=178051;dist=10154	rs2700947	0.001962839	INCREASE
chr7	36091147	intergenic	SEPT7,LOC101928618	dist=177907;dist=10298	rs2718005	0.001962839	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	0.002007182	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	0.002007182	INCREASE
chr11	72175110	intronic	STARD10	NA	rs481206	0.00204715	DECREASE
chr17	46276879	intergenic	WFIKK2,TOB1	dist=2171;dist=17707	rs9893135	0.002169912	INCREASE
chr15	61467142	intergenic	CA12,LOC102723344	dist=5780;dist=2340	rs35474601	0.002203505	DECREASE
chr9	76905965	intronic	OSTF1	NA	rs11144228	0.002276914	DECREASE
chr3	1.27E+08	ncRNA_exonic	FAM86JP	NA	rs13063122	0.002310662	DECREASE
chr16	82538204	intergenic	MLYCD,OSGIN1	dist=30916;dist=6124	rs4782862	0.002406649	DECREASE

**Table 13 continued**

chr12	1.15E+08	intergenic	MIR4472-2,LINC00173	dist=66155;dist=38949	rs1566930	0.002413074	INCREASE
chr8	1.02E+08	intergenic	FLJ42969,ZNF706	dist=44759;dist=76028	rs2441704	0.003600164	DECREASE
chr16	82538204	intergenic	MLYCD,OSGIN1	dist=30916;dist=6124	rs4782862	0.004888673	INCREASE
chr16	82538204	intergenic	MLYCD,OSGIN1	dist=30916;dist=6124	rs4782862	0.004888673	INCREASE
chr3	1.66E+08	intergenic	LINC01192,SI	dist=1331742;dist=343855	rs34741709	0.00528453	DECREASE
chr3	1.66E+08	intergenic	LINC01192,SI	dist=1331742;dist=343855	rs34741709	0.00528453	DECREASE
chr6	53437811	intergenic	ELOVL5,GCLC	dist=115875;dist=32288	rs593532	0.005296121	INCREASE
chr17	72795651	UTR5	41891	NA	rs138270732	0.005319381	DECREASE
chr7	7328048	intergenic	LOC101927354,COL28A1	dist=43823;dist=36721	rs1882600	0.005717452	DECREASE
chr7	7328048	intergenic	LOC101927354,COL28A1	dist=43823;dist=36721	rs1882600	0.005717452	DECREASE
chr5	1.32E+08	intronic	SLC22A5	NA	rs274567	0.006156115	DECREASE
chr5	1.32E+08	intronic	SLC22A5	NA	rs274567	0.006156115	DECREASE
chr8	99246585	intergenic	POP1,NIPAL2	dist=5340;dist=26978	rs1868993	0.006795569	DECREASE
chr8	99246585	intergenic	POP1,NIPAL2	dist=5340;dist=26978	rs1868993	0.006795569	DECREASE
chr17	78165175	downstream	WDR45B	NA	rs8071743	0.006931985	INCREASE
chr1	1.14E+08	UTR3	HIPK1	NA	rs10732635	0.007172273	DECREASE
chr1	1.14E+08	UTR3	HIPK1	NA	rs10732635	0.007172273	DECREASE
chr4	2908816	ncRNA_exonic	NOP14-AS1	NA	rs1263338	0.007333009	DECREASE
chr4	2908816	ncRNA_exonic	NOP14-AS1	NA	rs1263338	0.007333009	DECREASE
chr16	84050383	intergenic	MIR5093,GSE1	dist=152951;dist=152147	rs731957	0.007740416	INCREASE
chr16	84050383	intergenic	MIR5093,GSE1	dist=152951;dist=152147	rs731957	0.007740416	INCREASE
chr12	1.24E+08	intronic	SCARB1	NA	rs10773109	0.009679985	INCREASE
chr12	1.24E+08	intronic	SCARB1	NA	rs10773109	0.009679985	INCREASE
chr16	9049152	intergenic	USP7,C16orf72	dist=84310;dist=43886	rs113403549	0.009679985	DECREASE
chr16	9049152	intergenic	USP7,C16orf72	dist=84310;dist=43886	rs113403549	0.009679985	DECREASE
chr8	1.02E+08	intergenic	FLJ42969,ZNF706	dist=44759;dist=76028	rs2441704	0.009679985	DECREASE
chr8	1.02E+08	intergenic	FLJ42969,ZNF706	dist=44759;dist=76028	rs2441704	0.009679985	DECREASE
chr1	1.44E+08	upstream	PDZK1	NA	rs900347	0.009679985	INCREASE
chr1	1.44E+08	upstream	PDZK1	NA	rs900347	0.009679985	INCREASE

**Table 13 continued**

chr17	37203179	intergenic	JUP,LEPREL4	dist=6689;dist=8552	rs12453367	0.009786707	INCREASE
chr17	46217672	intergenic	MIR8059,WFIKK2	dist=16582;dist=49932	rs4793670	0.010569059	INCREASE
chr17	46217672	intergenic	MIR8059,WFIKK2	dist=16582;dist=49932	rs4793670	0.010569059	INCREASE
chr14	68482557	intronic	ACTN1	NA	rs2268982	0.011143964	DECREASE
chr17	72794878	intronic	41891	NA	rs1996631	0.014593517	INCREASE
chr11	1.01E+08	intergenic	LOC101054525,TRPC6	dist=248353;dist=43941	rs12785425	0.018701016	INCREASE
chr11	1.01E+08	intergenic	LOC101054525,TRPC6	dist=248353;dist=43941	rs12785425	0.018701016	INCREASE
chr11	1.01E+08	intergenic	LOC101054525,TRPC6	dist=248353;dist=43941	rs12785425	0.018701016	INCREASE
chr10	12232323	intronic	SEC61A2	NA	rs9329337	0.0199501	DECREASE
chr12	51650500	intergenic	KRT18,EIF4B	dist=17548;dist=35829	rs2682327	0.021292451	DECREASE
chr16	53015651	intergenic	IRX3,CRNDE	dist=137772;dist=494627	rs9940750	0.026543734	DECREASE
chr16	53015651	intergenic	IRX3,CRNDE	dist=137772;dist=494627	rs9940750	0.026543734	DECREASE
chr16	53015651	intergenic	IRX3,CRNDE	dist=137772;dist=494627	rs9940750	0.026543734	DECREASE
chr3	1.78E+08	intergenic	LINC01209,TBL1XR1	dist=135007;dist=68746	rs2862637	0.030321457	INCREASE
chr3	1.78E+08	intergenic	LINC01209,TBL1XR1	dist=135007;dist=68746	rs2862637	0.030321457	INCREASE

**Table 14. The list of regulatory SNVs in TAMR Cell line**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr8	1.29E+08	ncRNA_intronic	PVT1	NA	chr8:128992864	2.25786E-06	INCREASE
chrX	19391837	intronic	MAP3K15	NA	rs5909298	7.38457E-06	DECREASE
chr1	1.44E+08	UTR5	PDE4DIP	NA	rs1324349	7.77221E-06	INCREASE
chrX	19391837	intronic	MAP3K15	NA	rs5909298	7.9065E-06	DECREASE
chrX	19391837	intronic	MAP3K15	NA	rs5909298	7.9065E-06	DECREASE
chr12	25926409	intergenic	MIR4302,RASSF8-AS1	dist=8130;dist=72446	rs12368327	1.03041E-05	INCREASE
chrX	1.49E+08	ncRNA_intronic	LINC00894	NA	rs1080027	1.23128E-05	INCREASE
chr11	1.12E+08	intergenic	LOC387810,LOC101928847	dist=152013;dist=251464	rs2055936	1.23128E-05	INCREASE
chr3	1.99E+08	intergenic	LOC220729,KIAA0226	dist=31693;dist=11820	rs145563991	1.26161E-05	INCREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr5	4984128	intergenic	LOC101929153,LINC01020	dist=156150;dist=103344	rs621356	1.50767E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	1.88217E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	1.88217E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr7	1.55E+08	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr7	1.55E+08	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr7	1.55E+08	intergenic	SHH,LOC389602	dist=50726;dist=99633	rs34044649	2.0412E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	2.20437E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE

Table 14 continued

chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE
chr11	64982763	intergenic	MIR612,MALAT1	dist=14159;dist=39046	rs1626021	2.4987E-05	INCREASE
chr5	4984128	intergenic	LOC101929153,LINC01020	dist=156150;dist=103344	rs621356	2.98815E-05	INCREASE
chr5	4984128	intergenic	LOC101929153,LINC01020	dist=156150;dist=103344	rs621356	2.98815E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr11	44297292	intergenic	ALX4,CD82	dist=9000;dist=246425	rs7480769	3.28554E-05	DECREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr2	1.06E+08	intergenic	UXS1,PLGLA	dist=108734;dist=79041	rs3890642	3.62837E-05	DECREASE
chr2	1.06E+08	intergenic	UXS1,PLGLA	dist=108734;dist=79041	rs3890642	3.62837E-05	DECREASE
chr2	1.06E+08	intergenic	UXS1,PLGLA	dist=108734;dist=79041	rs3890642	3.62837E-05	DECREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chrX	1.49E+08	ncRNA_intronic	LINC00894	NA	rs1080027	4.01757E-05	INCREASE
chrX	1.49E+08	ncRNA_intronic	LINC00894	NA	rs1080027	4.01757E-05	INCREASE
chr2	2.22E+08	intergenic	MIR4268,EPHA4	dist=1451402;dist=60059	rs2011122	4.10495E-05	INCREASE
chr2	2.22E+08	intergenic	MIR4268,EPHA4	dist=1451402;dist=60059	rs2011122	4.10495E-05	INCREASE
chr2	2.22E+08	intergenic	MIR4268,EPHA4	dist=1451402;dist=60059	rs2011122	4.10495E-05	INCREASE
chr14	74515113	intergenic	PGF,EIF2B2	dist=22893;dist=24252	rs175005	4.27669E-05	INCREASE
chr8	18926781	intergenic	PSD3,LOC100128993	dist=11305;dist=158685	rs1426916	4.3212E-05	INCREASE
chr8	18926781	intergenic	PSD3,LOC100128993	dist=11305;dist=158685	rs1426916	4.3212E-05	INCREASE
chr8	18926781	intergenic	PSD3,LOC100128993	dist=11305;dist=158685	rs1426916	4.3212E-05	INCREASE
chr20	4090809	intronic	SMOX	NA	rs13040038	4.38072E-05	DECREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE

**Table 14 continued**

chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr5	17144822	intergenic	MYO10,LOC285696	dist=155437;dist=38315	rs79986080	4.73771E-05	INCREASE
chr2	2.24E+08	intronic	WDFY1	NA	rs538980	4.75328E-05	INCREASE
chr2	2.24E+08	intronic	WDFY1	NA	rs538980	4.75328E-05	INCREASE
chr2	2.24E+08	intronic	WDFY1	NA	rs538980	4.75328E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr4	89382877	intergenic	ABCG2,PPM1K	dist=11379;dist=14908	rs997630	5.03646E-05	DECREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr7	1.31E+08	intergenic	PODXL,LOC101928782	dist=69435;dist=284168	rs2971746	5.05513E-05	DECREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr16	3955317	UTR3	ADCY9	NA	rs1045475	5.43508E-05	INCREASE
chr16	3955317	UTR3	ADCY9	NA	rs1045475	5.43508E-05	INCREASE
chr16	3955317	UTR3	ADCY9	NA	rs1045475	5.43508E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.02636E-05	INCREASE
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE

Table 14 continued							
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE
chr8	19292641	intronic	SH2D4A	NA	rs2410611	6.18167E-05	INCREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	6.2727E-05	DECREASE
chr2	15788944	intergenic	LOC101926966,MYCNOS	dist=12408;dist=204894	rs880565	6.48179E-05	INCREASE
chr2	15788944	intergenic	LOC101926966,MYCNOS	dist=12408;dist=204894	rs880565	6.48179E-05	INCREASE
chr11	44697969	intergenic	CD82,TSPAN18	dist=100078;dist=44583	rs7950389	6.55787E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE
chr16	87569119	intronic	CBFA2T3	NA	rs4782488	6.99845E-05	DECREASE
chr16	87569119	intronic	CBFA2T3	NA	rs4782488	6.99845E-05	DECREASE
chr16	87569119	intronic	CBFA2T3	NA	rs4782488	6.99845E-05	DECREASE
chr1	2E+08	intronic	LGR6	NA	rs1318189	7.09531E-05	DECREASE
chr1	2E+08	intronic	LGR6	NA	rs1318189	7.09531E-05	DECREASE
chr12	25926409	intergenic	MIR4302,RASSF8-AS1	dist=8130;dist=72446	rs12368327	7.49599E-05	INCREASE



**Table 15. The list of regulatory SNVs in ZR75 Cell line.**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr1	2.1E+08	intergenic	LINC00467,RD3	dist=37815;dist=6172	rs77207115	2.17145E-05	INCREASE
chr1	2.1E+08	intergenic	LINC00467,RD3	dist=37815;dist=6172	rs77207115	2.17145E-05	INCREASE
chr1	2.1E+08	intergenic	LINC00467,RD3	dist=37815;dist=6172	rs77207115	2.17145E-05	INCREASE
chr16	317624	intronic	AXIN1	NA	rs10903014	2.53932E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	3.33543E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	3.33543E-05	DECREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	3.46458E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	22998700	intronic	USP31	NA	chr16:22998700	6.48179E-05	DECREASE
chr16	22998700	intronic	USP31	NA	chr16:22998700	6.48179E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr6	1.49E+08	intergenic	SASH1,UST	dist=81959;dist=113128	rs74400481	6.86397E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	7.22241E-05	DECREASE
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	8.15262E-05	DECREASE



Table 15 continued							
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	8.15262E-05	DECREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chr5	66546818	intergenic	CD180,LOC102467655	dist=18445;dist=974642	rs1705397	8.25737E-05	INCREASE
chr10	79402130	intergenic	DLG5-AS1,POLR3A	dist=42541;dist=2783	rs10160016	8.84467E-05	INCREASE
chr10	79402130	intergenic	DLG5-AS1,POLR3A	dist=42541;dist=2783	rs10160016	8.84467E-05	INCREASE
chr10	79402130	intergenic	DLG5-AS1,POLR3A	dist=42541;dist=2783	rs10160016	8.84467E-05	INCREASE
chr9	1.3E+08	exonic	PIP5KL1	NA	rs10760515	9.87132E-05	DECREASE
chr9	1.3E+08	exonic	PIP5KL1	NA	rs10760515	9.87132E-05	DECREASE
chr12	1.32E+08	intergenic	LOC101928416,FBRSL1	dist=134960;dist=24036	rs12309386	9.88261E-05	DECREASE
chr12	1.32E+08	intergenic	LOC101928416,FBRSL1	dist=134960;dist=24036	rs12309386	9.88261E-05	DECREASE
chr12	1.32E+08	intergenic	LOC101928416,FBRSL1	dist=134960;dist=24036	rs12309386	9.88261E-05	DECREASE
chr15	87794879	intergenic	LINC00925,RHCG	dist=52157;dist=20765	chr15:87794879	0.00010296	DECREASE
chr15	87794879	intergenic	LINC00925,RHCG	dist=52157;dist=20765	chr15:87794879	0.00010296	DECREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	0.000103704	INCREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	0.000103704	INCREASE
chrX	1.37E+08	intergenic	ZIC3,LINC00889	dist=465546;dist=577087	rs5931289	0.000110161	DECREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.00011304	INCREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.00011304	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	0.0001149	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	0.0001149	INCREASE
chr8	29632859	intergenic	DUSP4,LINC00589	dist=368673;dist=1835	rs10216706	0.000116641	INCREASE
chr8	29632859	intergenic	DUSP4,LINC00589	dist=368673;dist=1835	rs10216706	0.000116641	INCREASE
chr8	29632859	intergenic	DUSP4,LINC00589	dist=368673;dist=1835	rs10216706	0.000116641	INCREASE
chr15	55671170	upstream	GCOM1,MYZAP	NA	rs2641571	0.000119481	INCREASE
chr5	1.27E+08	intergenic	CTXN3,LINC01184	dist=98022;dist=264900	rs13182486	0.000130941	INCREASE
chr5	1.27E+08	intergenic	CTXN3,LINC01184	dist=98022;dist=264900	rs13182486	0.000130941	INCREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	0.000130941	INCREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	0.000130941	INCREASE

Table 15 continued							
chr16	4967077	intronic	SEC14L5	NA	rs2908649	0.000131768	INCREASE
chr4	1.84E+08	intronic	TENM3	NA	rs77905405	0.00013267	INCREASE
chr4	1.84E+08	intronic	TENM3	NA	rs77905405	0.00013267	INCREASE
chr14	93988897	upstream	SERPINA11	NA	rs1956720	0.000138172	INCREASE
chr14	93988897	upstream	SERPINA11	NA	rs1956720	0.000138172	INCREASE
chr14	93988897	upstream	SERPINA11	NA	rs1956720	0.000138172	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr5	66547543	intergenic	CD180,LOC102467655	dist=19170;dist=973917	rs1697138	0.000144987	INCREASE
chr15	99525863	intergenic	LRRK1,CHSY1	dist=98023;dist=7588	rs1982489	0.000144987	INCREASE
chr7	1.57E+08	exonic	PTPRN2	NA	rs1130502	0.00015549	DECREASE
chr7	1.57E+08	exonic	PTPRN2	NA	rs1130502	0.00015549	DECREASE
chr7	1.57E+08	exonic	PTPRN2	NA	rs1130502	0.00015549	DECREASE
chr4	1.75E+08	intergenic	HAND2-AS1,FBXO8	dist=675258;dist=19571	rs11946728	0.00015549	DECREASE
chr4	1.75E+08	intergenic	HAND2-AS1,FBXO8	dist=675258;dist=19571	rs11946728	0.00015549	DECREASE
chr4	1.75E+08	intergenic	HAND2-AS1,FBXO8	dist=675258;dist=19571	rs11946728	0.00015549	DECREASE
chr14	64503470	intronic	CHURC1-FNTB,RAB15	NA	rs8004904	0.000155556	INCREASE
chr14	64503470	intronic	CHURC1-FNTB,RAB15	NA	rs8004904	0.000155556	INCREASE
chr10	1.23E+08	intronic	FGFR2	NA	rs11599804	0.000158241	INCREASE
chr10	1.23E+08	intronic	FGFR2	NA	rs11599804	0.000158241	INCREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000162941	DECREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000162941	DECREASE
chr9	2459742	intergenic	SMARCA2,VLDLR-AS1	dist=276119;dist=65913	rs2150720	0.000165147	INCREASE
chr9	2459742	intergenic	SMARCA2,VLDLR-AS1	dist=276119;dist=65913	rs2150720	0.000165147	INCREASE
chr12	56403622	downstream	AGAP2	NA	rs238516	0.000165147	DECREASE
chr12	56403622	downstream	AGAP2	NA	rs238516	0.000165147	DECREASE
chr17	46278218	intergenic	WFIKK2,TOB1	dist=3510;dist=16368	rs8072476	0.000165147	DECREASE
chr17	46278218	intergenic	WFIKK2,TOB1	dist=3510;dist=16368	rs8072476	0.000165147	DECREASE

Table 15 continued							
chr20	54637885	UTR5	TFAP2C	NA	rs62208554	0.000167187	DECREASE
chr20	54637885	UTR5	TFAP2C	NA	rs62208554	0.000167187	DECREASE
chr16	57062847	intronic	NDRG4	NA	rs16960169	0.000169076	INCREASE
chr17	46217715	intergenic	MIR8059,WFIKKN2	dist=16625;dist=49889	rs12942595	0.000178475	DECREASE
chr17	46217715	intergenic	MIR8059,WFIKKN2	dist=16625;dist=49889	rs12942595	0.000178475	DECREASE
chr4	6748574	intronic	S100P	NA	rs3822263	0.000188278	DECREASE
chr18	9978324	intergenic	VAPA,LINC01254	dist=28306;dist=416806	rs29030	0.000191251	INCREASE
chr18	9978324	intergenic	VAPA,LINC01254	dist=28306;dist=416806	rs29030	0.000191251	INCREASE
chr18	9978324	intergenic	VAPA,LINC01254	dist=28306;dist=416806	rs29030	0.000191251	INCREASE
chr12	1.21E+08	intronic	TMEM120B	NA	rs4760112	0.000196412	INCREASE
chr12	1.21E+08	intronic	TMEM120B	NA	rs4760112	0.000196412	INCREASE
chr8	96775028	ncRNA_intronic	C8orf37-AS1	NA	rs7010069	0.000196706	INCREASE
chr8	96775028	ncRNA_intronic	C8orf37-AS1	NA	rs7010069	0.000196706	INCREASE
chr1	5871451	intronic	NPHP4	NA	rs9729880	0.000196736	INCREASE
chr6	12396408	intergenic	HIVEP1,EDN1	dist=123190;dist=2107	rs2859337	0.000201249	INCREASE
chr6	12396408	intergenic	HIVEP1,EDN1	dist=123190;dist=2107	rs2859337	0.000201249	INCREASE

**Table 16. The list of regulatory SNVs in good prognosis tumors**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr1	1.54E+08	intronic	PMF1,PMF1-BGLAP	NA	rs2475757	9.22878E-07	INCREASE
chr7	54699748	intergenic	VSTM2A-OT1,SEC61G	dist=92835;dist=87686	chr7:54699748	2.25786E-06	INCREASE
chr19	18253643	upstream	JUND,MIR3188	NA	rs41523455	7.56077E-06	DECREASE
chr1	1.44E+08	UTR5	PDE4DIP	NA	rs1324349	7.77221E-06	INCREASE
chr19	18253642	upstream	JUND,MIR3188	NA	rs41519246	7.77221E-06	DECREASE
chr2	1.33E+08	intergenic	ANKRD30BL,GPR39	dist=18939;dist=139666	rs62163790	7.77221E-06	INCREASE
chr1	2.31E+08	intronic	PCNXL2	NA	rs10797598	8.46441E-06	INCREASE
chr19	10597181	UTR5	SLC44A2	NA	rs139154013	8.46441E-06	DECREASE

**Table 16 continued**

chr1	43552151	exonic	TIE1	NA	rs3120276	1.06917E-05	INCREASE
chr1	2.13E+08	UTR5	PTPN14	NA	rs10864100	1.09518E-05	DECREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr11	1.12E+08	intergenic	LOC387810,LOC101928847	dist=152013;dist=251464	rs2055936	1.23128E-05	INCREASE
chr3	1.99E+08	intergenic	LOC220729,KIAA0226	dist=31693;dist=11820	rs145563991	1.26161E-05	INCREASE
chr1	2.29E+08	UTR5	AGT	NA	rs5050	1.31407E-05	DECREASE
chr1	2.29E+08	UTR5	AGT	NA	rs5050	1.31407E-05	DECREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr6	1.1E+08	upstream	CD164	NA	rs3757231	1.69288E-05	INCREASE
chr3	80041411	intergenic	ROBO1,GBE1	dist=141662;dist=1580129	rs4563372	1.99407E-05	DECREASE
chr3	80041411	intergenic	ROBO1,GBE1	dist=141662;dist=1580129	rs4563372	1.99407E-05	DECREASE
chr3	80041411	intergenic	ROBO1,GBE1	dist=141662;dist=1580129	rs4563372	1.99407E-05	DECREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr20	57464499	intergenic	EDN3,PHACTR3	dist=130057;dist=121460	rs271977	2.22362E-05	INCREASE
chr13	35535010	intronic	DCLK1	NA	rs17053384	2.45735E-05	INCREASE
chr13	35535010	intronic	DCLK1	NA	rs17053384	2.45735E-05	INCREASE
chr13	35535010	intronic	DCLK1	NA	rs17053384	2.45735E-05	INCREASE
chr16	317624	intronic	AXIN1	NA	rs10903014	2.53932E-05	INCREASE
chr1	1.2E+08	exonic	NOTCH2	NA	rs11810554	2.53932E-05	INCREASE
chr16	82707955	UTR5	MBTPS1	NA	rs7205403	2.53932E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE
chr14	99091387	intronic	CCDC85C	NA	rs10147748	3.13772E-05	DECREASE

Table 16 continued

chr1	2.33E+08	intergenic	LINC00184,LINC01132	dist=47691;dist=41572	rs12748673	3.2491E-05	DECREASE
chr1	2.33E+08	intergenic	LINC00184,LINC01132	dist=47691;dist=41572	rs12748673	3.2491E-05	DECREASE
chr1	2.33E+08	intergenic	LINC00184,LINC01132	dist=47691;dist=41572	rs12748673	3.2491E-05	DECREASE
chr11	44297292	intergenic	ALX4,CD82	dist=9000;dist=246425	rs7480769	3.28554E-05	DECREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chr1	2.07E+08	intergenic	PLXNA2,MIR205HG	dist=130558;dist=1053945	rs696983	3.64709E-05	INCREASE
chr19	50041017	upstream	PVRL2	NA	rs77241309	3.77088E-05	INCREASE
chr19	50041017	upstream	PVRL2	NA	rs77241309	3.77088E-05	INCREASE
chr22	24575178	ncRNA_intronic	MIR1302-1	NA	rs2859409	3.85968E-05	DECREASE
chr22	24575178	ncRNA_intronic	MIR1302-1	NA	rs2859409	3.85968E-05	DECREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr4	1.91E+08	intergenic	LINC01262,FRG1	dist=60136;dist=219198	rs5005522	4.27045E-05	INCREASE
chr4	1.91E+08	intergenic	LINC01262,FRG1	dist=60136;dist=219198	rs5005522	4.27045E-05	INCREASE
chr20	4090809	intronic	SMOX	NA	rs13040038	4.38072E-05	DECREASE
chr2	2.24E+08	intergenic	KCNE4,SCG2	dist=410440;dist=130863	rs1439926	4.42322E-05	DECREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	4.49834E-05	DECREASE
chr9	14396805	intergenic	NFIB,ZDHC21	dist=7823;dist=204264	rs1407836	4.51824E-05	DECREASE
chr19	61335967	intergenic	ZNF787,ZNF444	dist=11413;dist=8380	rs665082	4.73021E-05	INCREASE
chr19	61335967	intergenic	ZNF787,ZNF444	dist=11413;dist=8380	rs665082	4.73021E-05	INCREASE
chr19	61335967	intergenic	ZNF787,ZNF444	dist=11413;dist=8380	rs665082	4.73021E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr12	33030513	intergenic	PKP2,SYT10	dist=89466;dist=389102	rs7306057	4.93682E-05	INCREASE

**Table 16 continued**

chr12	6315224	intronic	TNFRSF1A	NA	rs34819705	5.05513E-05	INCREASE
chr12	6315224	intronic	TNFRSF1A	NA	rs34819705	5.05513E-05	INCREASE
chr16	33846021	intergenic	RNU6-76P,LINC00273	dist=375277;dist=22532	rs79467332	5.17572E-05	DECREASE
chr16	33846021	intergenic	RNU6-76P,LINC00273	dist=375277;dist=22532	rs79467332	5.17572E-05	DECREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr17	21489423	intergenic	C17orf51,FAM27L	dist=93889;dist=260074	rs79109536	5.44256E-05	DECREASE
chr17	21489423	intergenic	C17orf51,FAM27L	dist=93889;dist=260074	rs79109536	5.44256E-05	DECREASE
chr17	54818764	intronic	YPEL2	NA	rs8073731	5.44256E-05	DECREASE
chr17	54818764	intronic	YPEL2	NA	rs8073731	5.44256E-05	DECREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr12	33030513	intergenic	PKP2,SYT10	dist=89466;dist=389102	rs7306057	5.62199E-05	INCREASE
chr12	33030513	intergenic	PKP2,SYT10	dist=89466;dist=389102	rs7306057	5.62199E-05	INCREASE
chr16	69717250	ncRNA_intronic	HYDIN2	NA	rs79976803	5.63588E-05	INCREASE
chr16	69717250	ncRNA_intronic	HYDIN2	NA	rs79976803	5.63588E-05	INCREASE
chr16	69717250	ncRNA_intronic	HYDIN2	NA	rs79976803	5.63588E-05	INCREASE
chr14	1.05E+08	intergenic	TMEM121,MIR8071-2	dist=73557;dist=17357	rs4983455	5.69393E-05	DECREASE
chr14	1.05E+08	intergenic	TMEM121,MIR8071-2	dist=73557;dist=17357	rs4983455	5.69393E-05	DECREASE
chr22	24575178	ncRNA_intronic	MIR1302-1	NA	rs2859409	5.84418E-05	DECREASE
chr8	1.45E+08	intergenic	ZNF623,ZNF707	dist=7754;dist=23813	chr8:144814797	6.18167E-05	DECREASE
chr8	1.45E+08	intergenic	ZNF623,ZNF707	dist=7754;dist=23813	chr8:144814797	6.18167E-05	DECREASE
chr8	1.45E+08	intergenic	ZNF623,ZNF707	dist=7754;dist=23813	chr8:144814797	6.18167E-05	DECREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr18	97864	intergenic	LOC102723376,ROCK1P1	dist=91934;dist=1201	chr18:97864	6.53221E-05	INCREASE
chr18	97864	intergenic	LOC102723376,ROCK1P1	dist=91934;dist=1201	chr18:97864	6.53221E-05	INCREASE
chr18	97864	intergenic	LOC102723376,ROCK1P1	dist=91934;dist=1201	chr18:97864	6.53221E-05	INCREASE
chr7	62110563	intergenic	NONE,ZNF733P	dist=NONE;dist=278542	rs73697257	6.53221E-05	INCREASE

**Table 16 continued**

chr7	62110563	intergenic	NONE,ZNF733P	dist=NONE;dist=278542	rs73697257	6.53221E-05	INCREASE
chr7	62110563	intergenic	NONE,ZNF733P	dist=NONE;dist=278542	rs73697257	6.53221E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE
chr8	8165597	intergenic	FAM86B3P,SGK223	dist=25800;dist=47071	rs2955552	6.72671E-05	INCREASE

**Table 17. The list of regulatory SNVs in bad prognosis tumors**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr7	1034712	intronic	C7orf50	NA	rs115018039	0	DECREASE
chr8	28006454	intronic	ELP3	NA	rs2305452	0	INCREASE
chr11	87548211	UTR5	RAB38	NA	rs3812730	0	DECREASE
chr6	1.19E+08	UTR5	MCM9	NA	rs62422268	0	DECREASE
chr17	71096951	ncRNA_exonic	MYO15B	NA	rs117940210	9.22878E-07	INCREASE
chr1	1.54E+08	intronic	PMF1,PMF1-BGLAP	NA	rs2475757	9.22878E-07	INCREASE
chr19	18391161	UTR5	SSBP4	NA	rs28375303	9.22878E-07	DECREASE
chr16	704754	upstream	METR1	NA	rs3809666	9.22878E-07	DECREASE
chr10	1.21E+08	upstream	GRK5	NA	rs10886423	1.03706E-06	DECREASE
chr21	33883005	upstream;downstream	DONSON;CRYZL1	NA	rs115794175	1.03706E-06	INCREASE
chr5	1.49E+08	UTR5	ARHGEF37	NA	rs35436322	1.03706E-06	DECREASE
chr12	6354970	UTR5	LTBR,SCNN1A	NA	rs10849447	2.25786E-06	INCREASE
chr19	19748174	ncRNA_exonic	LINC00663	NA	rs248944	2.25786E-06	DECREASE
chr8	8403033	intergenic	SGK223,CLDN23	dist=126279;dist=194043	rs56037578	2.79491E-06	INCREASE
chr10	1.19E+08	intronic	KIAA1598	NA	rs6585423	2.79491E-06	DECREASE
chr5	1.12E+08	intronic	EPB41L4A	NA	rs10478076	3.27257E-06	INCREASE
chr19	19518590	downstream	CILP2	NA	rs73924805	4.14822E-06	INCREASE
chr1	1.64E+08	intergenic	LOC400794,MGST3	dist=16097;dist=32672	rs10158373	5.38308E-06	DECREASE
chr19	4322431	intronic	SH3GL1	NA	rs104964	5.38308E-06	DECREASE



Table 17 continued

chr10	1.05E+08	intronic	SH3PXD2A	NA	rs7910092	5.38308E-06	INCREASE
chr1	1.14E+08	intronic	HIPK1	NA	rs1217441	7.56077E-06	INCREASE
chr10	97406097	intronic	ALDH18A1	NA	rs2225892	7.56077E-06	DECREASE
chr2	1.77E+08	ncRNA_splicing	LOC102724224	NR_110599:exon1:c.255+2C>T	rs2969356	7.56077E-06	DECREASE
chr19	18253643	upstream	JUND,MIR3188	NA	rs41523455	7.56077E-06	DECREASE
chr1	1.44E+08	UTR5	PDE4DIP	NA	rs1324349	7.77221E-06	INCREASE
chr19	18253642	upstream	JUND,MIR3188	NA	rs41519246	7.77221E-06	DECREASE
chr2	1.33E+08	intergenic	ANKRD30BL,GPR39	dist=18939;dist=139666	rs62163790	7.77221E-06	INCREASE
chr6	13982218	intergenic	MCUR1,RNF182	dist=59447;dist=50438	rs1204243	8.38474E-06	INCREASE
chr3	72468930	intergenic	LINC00870,RYBP	dist=162750;dist=37504	rs4677147	8.38474E-06	DECREASE
chr19	16443937	upstream	EPS15L1	NA	rs2885706	8.46441E-06	DECREASE
chr5	1.12E+08	intronic	EPB41L4A	NA	rs10478076	9.60381E-06	INCREASE
chr5	1.12E+08	intronic	EPB41L4A	NA	rs10478076	9.60381E-06	INCREASE
chr12	25926409	intergenic	MIR4302,RASSF8-AS1	dist=8130;dist=72446	rs12368327	1.03041E-05	INCREASE
chr4	1782841	downstream	LETM1	NA	rs10014663	1.06917E-05	DECREASE
chr8	1898800	intergenic	ARHGEF10,KBTBD11	dist=4586;dist=10651	rs17757154	1.06917E-05	DECREASE
chr6	34630750	intronic	SPDEF	NA	rs3798542	1.06917E-05	DECREASE
chr10	1.28E+08	upstream	FANK1	NA	rs78422237	1.06917E-05	DECREASE
chr16	30314770	intronic	ZNF48	NA	rs13333093	1.09518E-05	DECREASE
chr8	1.44E+08	UTR5	THEM6	NA	rs2280877	1.09518E-05	DECREASE
chr14	20387049	intergenic	RNASE1,RNASE3	dist=46173;dist=42353	rs28419520	1.09518E-05	DECREASE
chr10	94821513	intergenic	CYP26C1,CYP26A1	dist=3069;dist=1709	rs68040629	1.09518E-05	INCREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr8	98845752	intergenic	MTDH,LAPTM4B	dist=34088;dist=11233	rs7827538	1.14577E-05	DECREASE
chr11	1.12E+08	intergenic	LOC387810,LOC101928847	dist=152013;dist=251464	rs2055936	1.23128E-05	INCREASE
chr5	1.12E+08	UTR5	APC	NA	rs138386816	1.26161E-05	DECREASE
chr18	9604096	intronic	PPP4R1	NA	rs71360847	1.26161E-05	INCREASE
chr2	10009824	intronic	GRHL1	NA	rs73913927	1.26161E-05	DECREASE
chr3	75801255	intergenic	FLJ20518,LINC00960	dist=2197;dist=2867	rs189438805	1.36538E-05	DECREASE



**Table 17 continued**

chr1	26369042	UTR5	ZNF593	NA	rs2232648	1.36538E-05	DECREASE
chr19	1016044	exonic	ABCA7	NA	rs4147935	1.36538E-05	DECREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr6	77538132	intergenic	IMPG1,HTR1B	dist=699017;dist=690535	rs67307959	1.36538E-05	DECREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr17	45964961	upstream	EPN3	NA	rs12953066	1.51215E-05	DECREASE
chr4	39734053	ncRNA_intronic	LOC344967	NA	rs1706025	1.61492E-05	INCREASE
chr6	1.1E+08	upstream	CD164	NA	rs3757231	1.69288E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr19	16018435	intergenic	LINC00905,TPM4	dist=4490;dist=20882	rs4808432	1.93066E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr15	97136484	intronic	IGF1R	NA	rs62022087	2.02609E-05	INCREASE
chr16	317624	intronic	AXIN1	NA	rs10903014	2.53932E-05	INCREASE
chr14	1.04E+08	intronic	ADSSL1	NA	rs117084961	2.53932E-05	INCREASE
chr1	1.2E+08	exonic	NOTCH2	NA	rs11810554	2.53932E-05	INCREASE
chr11	2377970	intergenic	CD81,TSSC4	dist=2745;dist=2129	rs800351	2.53932E-05	INCREASE
chr20	31401767	intergenic	BPIFB1,CDK5RAP1	dist=40422;dist=8539	rs293709	2.89476E-05	INCREASE
chr20	31401767	intergenic	BPIFB1,CDK5RAP1	dist=40422;dist=8539	rs293709	2.89476E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr6	52573007	intergenic	TRAM2-AS1,LOC730101	dist=16257;dist=64151	rs10948704	3.13527E-05	INCREASE
chr6	52573007	intergenic	TRAM2-AS1,LOC730101	dist=16257;dist=64151	rs10948704	3.13527E-05	INCREASE
chr6	52573007	intergenic	TRAM2-AS1,LOC730101	dist=16257;dist=64151	rs10948704	3.13527E-05	INCREASE

**Table 17 continued**

chr12	45051495	intronic	SLC38A2	NA	rs10880961	3.28554E-05	DECREASE
chr2	730277	intergenic	TMEM18,LINC01115	dist=62838;dist=39560	rs112665529	3.28554E-05	DECREASE
chr5	77692527	intronic	SCAMP1	NA	rs4634329	3.28554E-05	INCREASE
chr9	38059463	upstream	SHB	NA	rs79230237	3.28554E-05	INCREASE
chr20	62051086	intronic	UCKL1	NA	rs817310	3.28554E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	3.33543E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	3.33543E-05	DECREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr13	99376302	ncRNA_intronic	LOC101927437	NA	rs9300575	3.37009E-05	INCREASE
chr10	1.21E+08	upstream	RGS10	NA	rs10787978	3.38577E-05	INCREASE
chr12	7199694	intronic	CLSTN3	NA	rs3782920	3.38577E-05	INCREASE
chr20	24878801	intronic	CST7	NA	rs227654	3.41023E-05	INCREASE
chr20	24878801	intronic	CST7	NA	rs227654	3.41023E-05	INCREASE
chr2	1.83E+08	intronic	PPP1R1C	NA	rs1515891	3.46458E-05	INCREASE
chr16	83472565	intronic	CRISPLD2	NA	rs2172622	3.46458E-05	INCREASE
chr9	68116030	upstream	LOC100132352	NA	rs78173369	3.62837E-05	DECREASE
chr9	68116030	upstream	LOC100132352	NA	rs78173369	3.62837E-05	DECREASE
chr19	49280118	intronic	ZNF284	NA	rs55771992	3.65661E-05	DECREASE
chr19	49280118	intronic	ZNF284	NA	rs55771992	3.65661E-05	DECREASE
chr2	1E+08	intergenic	LINC01104,LONRF2	dist=1756;dist=20051	rs2309812	3.78484E-05	DECREASE
chr2	1.6E+08	intronic	CD302,LY75-CD302	NA	rs2556106	3.78484E-05	DECREASE

**Table 18. The list of regulatory SNVs in metastatic tumors**

CHR	POS	ANNOTATION	GENE	DISTANCE FROM GENE	SNP ID	ADJ PVALUE FDR	SNP AFFECT ON BINDING
chr1	1.54E+08	intronic	PMF1,PMF1-BGLAP	NA	rs2475757	9.22878E-07	INCREASE
chr21	42122206	intronic	PRDM15	NA	rs62214691	2.79491E-06	DECREASE
chr10	1.19E+08	intronic	KIAA1598	NA	rs6585423	2.79491E-06	DECREASE
chr1	1.44E+08	UTR5	PDE4DIP	NA	rs1324349	7.77221E-06	INCREASE
chr3	72468930	intergenic	LINC00870,RYPB	dist=162750;dist=37504	rs4677147	8.38474E-06	DECREASE
chr2	42091558	intergenic	C2orf91,PKDCC	dist=57111;dist=37107	rs11887745	1.08813E-05	DECREASE
chr2	42091558	intergenic	C2orf91,PKDCC	dist=57111;dist=37107	rs11887745	1.08813E-05	DECREASE
chr2	42091558	intergenic	C2orf91,PKDCC	dist=57111;dist=37107	rs11887745	1.08813E-05	DECREASE
chr7	1.49E+08	ncRNA_intronic	ATP6V0E2-AS1	NA	rs185874861	1.09518E-05	DECREASE
chr14	20387049	intergenic	RNASE1,RNASE3	dist=46173;dist=42353	rs28419520	1.09518E-05	DECREASE
chr11	1.12E+08	intergenic	LOC387810,LOC101928847	dist=152013;dist=251464	rs2055936	1.23128E-05	INCREASE
chr19	1016044	exonic	ABCA7	NA	rs4147935	1.36538E-05	DECREASE
chr16	78615542	intergenic	LOC101928248,LOC102724084	dist=222801;dist=131814	rs4889067	1.36538E-05	DECREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr12	28619812	intergenic	CCDC91,FAR2	dist=25446;dist=573391	rs7315906	1.36859E-05	INCREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.37513E-05	INCREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.37513E-05	INCREASE
chr4	7826420	ncRNA_exonic	AFAP1-AS1	NA	rs10026941	1.45084E-05	INCREASE
chr4	7826420	ncRNA_exonic	AFAP1-AS1	NA	rs10026941	1.45084E-05	INCREASE
chr4	7826420	ncRNA_exonic	AFAP1-AS1	NA	rs10026941	1.45084E-05	INCREASE
chr17	45964961	upstream	EPN3	NA	rs12953066	1.51215E-05	DECREASE
chr17	43481157	intronic	NFE2L1	NA	rs73327306	1.51215E-05	DECREASE
chr21	26264932	intronic	APP	NA	chr21:26264932	1.63219E-05	INCREASE
chr21	26264932	intronic	APP	NA	chr21:26264932	1.63219E-05	INCREASE
chr6	1.21E+08	intergenic	LOC285762,TBC1D32	dist=650496;dist=937677	rs79423570	1.72702E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE

Table 18 continued

chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr10	1.26E+08	intergenic	CHST15,OAT	dist=143993;dist=88756	rs10794194	1.87026E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	1.88217E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	1.88217E-05	INCREASE
chr22	41003755	intergenic	OGFRP1,LINC01315	dist=2943;dist=86595	rs134885	2.05954E-05	INCREASE
chr22	41003755	intergenic	OGFRP1,LINC01315	dist=2943;dist=86595	rs134885	2.05954E-05	INCREASE
chr22	41003755	intergenic	OGFRP1,LINC01315	dist=2943;dist=86595	rs134885	2.05954E-05	INCREASE
chr2	2.24E+08	intergenic	ACSL3,KCNE4	dist=33672;dist=74857	rs34356450	2.20437E-05	INCREASE
chr21	26264932	intronic	APP	NA	chr21:26264932	2.57641E-05	INCREASE
chr20	31401767	intergenic	BPIFB1,CDK5RAP1	dist=40422;dist=8539	rs293709	2.89476E-05	INCREASE
chr20	31401767	intergenic	BPIFB1,CDK5RAP1	dist=40422;dist=8539	rs293709	2.89476E-05	INCREASE
chrX	53730090	intronic	HUWE1	NA	rs7877957	3.02431E-05	INCREASE
chr19	16039250	upstream	TPM4	NA	rs36048257	3.28554E-05	DECREASE
chr11	44297292	intergenic	ALX4,CD82	dist=9000;dist=246425	rs7480769	3.28554E-05	DECREASE
chr12	7199694	intronic	CLSTN3	NA	rs3782920	3.38577E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr11	20014669	intronic	NAV2	NA	rs10741810	3.64524E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr12	54839276	intronic	MYL6	NA	rs35436573	3.77734E-05	INCREASE
chr20	61603720	intergenic	EEF1A2,PPDPF	dist=2608;dist=18857	rs880447	3.79596E-05	INCREASE
chr20	61603720	intergenic	EEF1A2,PPDPF	dist=2608;dist=18857	rs880447	3.79596E-05	INCREASE
chr20	61603720	intergenic	EEF1A2,PPDPF	dist=2608;dist=18857	rs880447	3.79596E-05	INCREASE
chr1	1.15E+08	intronic	SYCP1	NA	rs74116188	3.81718E-05	INCREASE
chr1	1.15E+08	intronic	SYCP1	NA	rs74116188	3.81718E-05	INCREASE
chr1	1.15E+08	intronic	SYCP1	NA	rs74116188	3.81718E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE
chr21	45009292	intergenic	TSPEAR,UBE2G2	dist=53369;dist=3631	rs658657	4.10495E-05	INCREASE

**Table 18 continued**

chr1	22340802	intronic	WNT4	NA	rs3820282	4.22013E-05	INCREASE
chr2	69479142	intronic	NFU1	NA	rs7580642	4.3212E-05	INCREASE
chr2	69479142	intronic	NFU1	NA	rs7580642	4.3212E-05	INCREASE
chr2	69479142	intronic	NFU1	NA	rs7580642	4.3212E-05	INCREASE
chr9	1.39E+08	intronic	NOXA1	NA	rs11497278	4.38072E-05	INCREASE
chr21	40876587	intronic	DSCAM	NA	rs74811265	4.38313E-05	INCREASE
chr21	40876587	intronic	DSCAM	NA	rs74811265	4.38313E-05	INCREASE
chr21	40876587	intronic	DSCAM	NA	rs74811265	4.38313E-05	INCREASE
chr1	93232394	intergenic	FAM69A,MTF2	dist=32727;dist=84986	rs4240963	4.42322E-05	INCREASE
chr9	14396805	intergenic	NFIB,ZDHHC21	dist=7823;dist=204264	rs1407836	4.51824E-05	DECREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr9	82484228	intergenic	LOC101927477,TLE1	dist=644938;dist=904190	rs1412283	4.73021E-05	INCREASE
chr16	10520159	intergenic	ATF7IP2,EMP2	dist=35163;dist=9621	rs4780936	4.73021E-05	DECREASE
chr16	10520159	intergenic	ATF7IP2,EMP2	dist=35163;dist=9621	rs4780936	4.73021E-05	DECREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr6	51986013	intronic	PKHD1	NA	rs1896972	4.87364E-05	INCREASE
chr15	61839069	intronic	HERC1	NA	rs7167066	4.87364E-05	INCREASE
chr15	61839069	intronic	HERC1	NA	rs7167066	4.87364E-05	INCREASE
chr1	1.87E+08	intergenic	PLA2G4A,BRINP3	dist=1443628;dist=1665056	rs1472003	5.17869E-05	INCREASE
chr9	1.36E+08	intergenic	RXRA,COL5A1	dist=9168;dist=192051	rs11103603	5.44256E-05	INCREASE
chr9	1.36E+08	intergenic	RXRA,COL5A1	dist=9168;dist=192051	rs11103603	5.44256E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr16	4967077	intronic	SEC14L5	NA	rs2908649	5.49688E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr2	10384622	intronic	HPCAL1	NA	rs2014889	5.62199E-05	INCREASE
chr9	1.36E+08	intergenic	RXRA,COL5A1	dist=9168;dist=192051	rs11103603	5.63588E-05	INCREASE
chr11	69377341	intergenic	FGF3,LOC101928443	dist=34212;dist=202643	rs7395799	5.63588E-05	INCREASE

**Table 18 continued**

chr11	69377341	intergenic	FGF3,LOC101928443	dist=34212;dist=202643	rs7395799	5.63588E-05	INCREASE
chr11	69377341	intergenic	FGF3,LOC101928443	dist=34212;dist=202643	rs7395799	5.63588E-05	INCREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	5.72577E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	5.9582E-05	DECREASE
chr3	1.31E+08	intergenic	PLXND1,TMCC1	dist=16877;dist=24176	rs11718169	5.9582E-05	DECREASE
chr20	31401767	intergenic	BPIFB1,CDK5RAP1	dist=40422;dist=8539	rs293709	6.18462E-05	INCREASE
chr16	73721127	intergenic	LDHD,ZFP1	dist=12956;dist=18795	rs12448032	6.2727E-05	DECREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr14	49135467	UTR5	LRR1	NA	rs2281836	6.2727E-05	INCREASE
chr7	30995158	intergenic	GHRHR,ADCYAP1R1	dist=9487;dist=63443	rs11760522	6.90096E-05	INCREASE
chr7	30995158	intergenic	GHRHR,ADCYAP1R1	dist=9487;dist=63443	rs11760522	6.90096E-05	INCREASE
chr7	30995158	intergenic	GHRHR,ADCYAP1R1	dist=9487;dist=63443	rs11760522	6.90096E-05	INCREASE

**Table 19. The allele frequency of top RegSNVs in ER binding sites with sufficient coverage**

RegSNP	Target gene	Binding impact	Coverage in binding site	Mut reads	Frequency (%)	Sample	Reference
rs36208869	GSTM1	INCREASING	22	22	100	Tumor	(4)
rs1412825_LRRIQ3	LRRIQ3	INCREASING	11	11	100	Tumor	(4)
rs10956142_ANXA13	ANXA13	INCREASING	55	55	100	Tumor	(4)
rs2939587_TM2D3	TM2D3	INCREASING	10	10	100	Tumor	(4)
rs1291363_HTR7P1	HTR7P1	INCREASING	14	14	100	Tumor	(4)
rs4418583_LDLRAP1	LDLRAP1	INCREASING	27	23	85.2	Tumor, ZR75 cell line	(4)

**Table 20. Sequence of targeted amplification primers.**

<b>Mutation</b>	<b>Forward primer</b>	<b>Reverse primer</b>	<b>Amplicon size</b>
<i>ESRI</i> K303R	GCCCGCTCATGATCAAACG	CGGCCGTCAGGGACAAG	57
<i>ESRI</i> S463P	GCTTCTCTCTCTCACTCTCTCT	AGGACTCGGTGGATATGGT	101
<i>ESRI</i> Y537C			
<i>ESRI</i> Y537N	CAAAGGCATGGAGCATCTGTA	TGAAGTAGAGCCCGCAGT	169
<i>ESRI</i> Y537S			
<i>ESRI</i> D538G			

**Table 21. Sequence of ddPCR primer and probes.**

<b>Mutation</b>	<b>Forward primer</b>	<b>Reverse primer</b>	<b>Mutant Probe</b>	<b>WT probe</b>	<b>Fluorescence</b>
<i>ESR1</i> K303R	GCCCGCTCATGATCAAACG	CGGCCGTCAGGGACAAG	CAGGCTGTTCTTCTTAG	CAGGCTGTTCTTCTTAG	FAM/VIC
<i>ESR1</i> S463P	CTCTGCGCATTTCAGGAGTGT A	CGGTGGATATGGTCCTTCTC TTC	CATTCTGCCCAGCACC	CACATTCTGTCCAGCACC	FAM/VIC
<i>ESR1</i> Y537C	CAGCATGAAGTGCAAGAAC GT	TGGGCGTCCAGCATCTC	CCCCTCTGTGACCTG	TGCCCCCTCTATGACCTG	FAM/VIC
<i>ESR1</i> Y537N	CTGTACAGCATGAAGTGCA AGAAC	TGGGCGTCCAGCATCTC	TGCCCCCTCAATGAC	TGGTGCCCCCTCTATGAC	FAM/VIC
<i>ESR1</i> Y537S	CAGCATGAAGTGCAAGAAC GT	TGGGCGTCCAGCATCTC	CCCTCTCTGACCTGC	CCCCTCTATGACCTGC	FAM/VIC
<i>ESR1</i> D538G	GCATGAAGTGCAAGAACGT G	AAGTGGCTTTGGTCCGTCT	TCTATGGCCTGCTGCTGGAG ATGCT	TCTATGACCTGCTGCTGGAG ATGCT	HEX/FAM



**Table 22. Cellularity and location of bone metastases.**

<b>ID</b>	<b>Tumor Cellularity</b>	<b>Bone site</b>
BM01	80	Lt. distal femur (knee)
BM02	N/A	Lt. Proximal humerus
BM03	70	Rt. Pelvis
BM04	10-20	Lt. Femur
BM06	70-80	Lt. Proximal femur (hip)
BM07	30	Lt. Distal humerus
BM08	60	Lt. Proximal femur (hip)
BM09	10	Rt prox humerus
BM10	>90	Rt Femur, L humerus
BM11	80	Rt femur (hip)
BM12	70	Rt humerus
BM14	<5	Lt femur

**Table 23. The sequence of sgRNA and oligos used to generate *ESR1* mutant cell lines**

Name	Length (bp)	Sequence
sgRNA targeting <i>ESR1</i> gene	20	TCTCCAGCAGCAGGTCATAG
Oligo for Y537S	70	GCGGTGGGCGTCCAGCATCTCCAGCAGCAGG TCAGAGAGGGGCACCACGTTCTTGCACTTCATGCTGTAC
Oligo for D538G	70	GTAGGCGGTGGGCGTCCAGCATCTCCAGCAGCAG GCCATAGAGGGGCACCACGTTCTTGCACTTCATGCT

**Table 24. Name of the novel mutant ER target genes shared between T47D and MCF7**

Gene	Overlap
BBOX1	MCF7-Y537S and T47D-Y537S
TRANK1	MCF7-Y537S and T47D-Y537S
SPRR1B	MCF7-Y537S and T47D-Y537S
ITGAM	MCF7-Y537S and T47D-Y537S
PYDC1	MCF7-Y537S and T47D-Y537S
AHNAK2	MCF7-Y537S and T47D-Y537S
KRT10	MCF7-Y537S and T47D-Y537S
S100A2	MCF7-Y537S and T47D-Y537S
CDSN	MCF7-Y537S and T47D-Y537S
UBD	MCF7-Y537S and T47D-Y537S
AC132217.4	MCF7-Y537S and T47D-Y537S
FKBP1B	MCF7-D538G and T47D-D538G

## BIBLIOGRAPHY

1. Green, S. and Chambon, P. (1988) Nuclear receptors enhance our understanding of transcription regulation. *Trends in genetics : TIG*, **4**, 309-314.
2. Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H. *et al.* (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58-64.
3. Joseph, R., Orlov, Y.L., Huss, M., Sun, W., Kong, S.L., Ukil, L., Pan, Y.F., Li, G., Lim, M., Thomsen, J.S. *et al.* (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Molecular systems biology*, **6**, 456.
4. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R. *et al.* (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, **481**, 389-393.
5. Schmidt, D., Schwalie, P.C., Ross-Innes, C.S., Hurtado, A., Brown, G.D., Carroll, J.S., Flicek, P. and Odom, D.T. (2010) A CTCF-independent role for cohesin in tissue-specific transcription. *Genome research*, **20**, 578-588.
6. Musgrove, E.A. and Sutherland, R.L. (2009) Biological determinants of endocrine resistance in breast cancer. *Nature reviews. Cancer*, **9**, 631-643.
7. Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F. *et al.* (2006) Genome-wide analysis of estrogen receptor binding sites. *Nature genetics*, **38**, 1289-1297.
8. Porter, W., Saville, B., Hoivik, D. and Safe, S. (1997) Functional synergy between the transcription factor Sp1 and the estrogen receptor. *Molecular endocrinology*, **11**, 1569-1580.
9. Weisz, A. and Rosales, R. (1990) Identification of an estrogen response element upstream of the human c-fos gene that binds the estrogen receptor and the AP-1 transcription factor. *Nucleic acids research*, **18**, 5097-5106.
10. Mohammed, H., Russell, I.A., Stark, R., Rueda, O.M., Hickey, T.E., Tarulli, G.A., Serandour, A.A., Birrell, S.N., Bruna, A., Saadi, A. *et al.* (2015) Progesterone receptor modulates ERalpha action in breast cancer. *Nature*, **523**, 313-317.
11. Losel, R.M., Falkenstein, E., Feuring, M., Schultz, A., Tillmann, H.C., Rossol-Haseroth, K. and Wehling, M. (2003) Nongenomic steroid action: controversies, questions, and answers. *Physiological reviews*, **83**, 965-1016.
12. Figtree, G.A., McDonald, D., Watkins, H. and Channon, K.M. (2003) Truncated estrogen receptor alpha 46-kDa isoform in human endothelial cells: relationship to acute activation of nitric oxide synthase. *Circulation*, **107**, 120-126.

13. Li, L., Haynes, M.P. and Bender, J.R. (2003) Plasma membrane localization and function of the estrogen receptor alpha variant (ER46) in human endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4807-4812.
14. Osborne, C.K. and Schiff, R. (2005) Estrogen-receptor biology: continuing progress and therapeutic implications. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **23**, 1616-1622.
15. Song, R.X., McPherson, R.A., Adam, L., Bao, Y., Shupnik, M., Kumar, R. and Santen, R.J. (2002) Linkage of rapid estrogen action to MAPK activation by ERalpha-Shc association and Shc pathway activation. *Molecular endocrinology*, **16**, 116-127.
16. Deroo, B.J. and Korach, K.S. (2006) Estrogen receptors and human disease. *The Journal of clinical investigation*, **116**, 561-570.
17. Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., van't Veer, L.J. and Perou, C.M. (2006) Concordance among gene-expression-based predictors for breast cancer. *The New England journal of medicine*, **355**, 560-569.
18. Howlader, N., Altekruse, S.F., Li, C.I., Chen, V.W., Clarke, C.A., Ries, L.A. and Cronin, K.A. (2014) US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *Journal of the National Cancer Institute*, **106**.
19. Voduc, K.D., Cheang, M.C., Tyldesley, S., Gelmon, K., Nielsen, T.O. and Kennecke, H. (2010) Breast cancer subtypes and the risk of local and regional relapse. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **28**, 1684-1691.
20. Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. *et al.* (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 10869-10874.
21. Massarweh, S. and Schiff, R. (2006) Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk. *Endocrine-related cancer*, **13 Suppl 1**, S15-24.
22. (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet*, **351**, 1451-1467.
23. Jaiyesimi, I.A., Buzdar, A.U., Decker, D.A. and Hortobagyi, G.N. (1995) Use of tamoxifen for breast cancer: twenty-eight years later. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **13**, 513-529.
24. Howell, A., Osborne, C.K., Morris, C. and Wakeling, A.E. (2000) ICI 182,780 (Faslodex): development of a novel, "pure" antiestrogen. *Cancer*, **89**, 817-825.
25. Osborne, C.K. (1999) Aromatase inhibitors in relation to other forms of endocrine therapy for breast cancer. *Endocrine-related cancer*, **6**, 271-276.
26. Bonnetterre, J., Thurlimann, B., Robertson, J.F., Krzakowski, M., Mauriac, L., Koralewski, P., Vergote, I., Webster, A., Steinberg, M. and von Euler, M. (2000) Anastrozole versus tamoxifen as first-line therapy for advanced breast cancer in 668 postmenopausal women: results of the Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **18**, 3748-3757.
27. Mouridsen, H., Gershanovich, M., Sun, Y., Perez-Carrion, R., Boni, C., Monnier, A., Apffelstaedt, J., Smith, R., Sleeboom, H.P., Janicke, F. *et al.* (2001) Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced

- breast cancer: results of a phase III study of the International Letrozole Breast Cancer Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **19**, 2596-2606.
28. Nabholz, J.M., Buzdar, A., Pollak, M., Harwin, W., Burton, G., Mangalik, A., Steinberg, M., Webster, A. and von Euler, M. (2000) Anastrozole is superior to tamoxifen as first-line therapy for advanced breast cancer in postmenopausal women: results of a North American multicenter randomized trial. Arimidex Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **18**, 3758-3767.
  29. Howell, A., Cuzick, J., Baum, M., Buzdar, A., Dowsett, M., Forbes, J.F., Hocht-Boes, G., Houghton, J., Locker, G.Y., Tobias, J.S. *et al.* (2005) Results of the ATAC (Arimidex, Tamoxifen, Alone or in Combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet*, **365**, 60-62.
  30. Howell, A., Robertson, J.F., Quaresma Albano, J., Aschermannova, A., Mauriac, L., Kleeberg, U.R., Vergote, I., Erikstein, B., Webster, A. and Morris, C. (2002) Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **20**, 3396-3403.
  31. Mehta, R.S., Barlow, W.E., Albain, K.S., Vandenberg, T.A., Dakhil, S.R., Tirumali, N.R., Lew, D.L., Hayes, D.F., Gralow, J.R., Livingston, R.B. *et al.* (2012) Combination anastrozole and fulvestrant in metastatic breast cancer. *The New England journal of medicine*, **367**, 435-444.
  32. Osborne, C.K., Pippin, J., Jones, S.E., Parker, L.M., Ellis, M., Come, S., Gertler, S.Z., May, J.T., Burton, G., Dimery, I. *et al.* (2002) Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **20**, 3386-3395.
  33. Droog, M., Beelen, K., Linn, S. and Zwart, W. (2013) Tamoxifen resistance: from bench to bedside. *European journal of pharmacology*, **717**, 47-57.
  34. Osborne, C.K., Bardou, V., Hopp, T.A., Chamness, G.C., Hilsenbeck, S.G., Fuqua, S.A., Wong, J., Allred, D.C., Clark, G.M. and Schiff, R. (2003) Role of the estrogen receptor coactivator AIB1 (SRC-3) and HER-2/neu in tamoxifen resistance in breast cancer. *Journal of the National Cancer Institute*, **95**, 353-361.
  35. deGraffenried, L.A., Friedrichs, W.E., Russell, D.H., Donzis, E.J., Middleton, A.K., Silva, J.M., Roth, R.A. and Hidalgo, M. (2004) Inhibition of mTOR activity restores tamoxifen response in breast cancer cells with aberrant Akt Activity. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **10**, 8059-8067.
  36. Britton, D.J., Hutcheson, I.R., Knowlden, J.M., Barrow, D., Giles, M., McClelland, R.A., Gee, J.M. and Nicholson, R.I. (2006) Bidirectional cross talk between ERalpha and EGFR signalling pathways regulates tamoxifen-resistant growth. *Breast cancer research and treatment*, **96**, 131-146.
  37. Drury, S.C., Detre, S., Leary, A., Salter, J., Reis-Filho, J., Barbashina, V., Marchio, C., Lopez-Knowles, E., Ghazoui, Z., Habben, K. *et al.* (2011) Changes in breast cancer biomarkers in the IGF1R/PI3K pathway in recurrent breast cancer after tamoxifen treatment. *Endocrine-related cancer*, **18**, 565-577.

38. Parisot, J.P., Hu, X.F., DeLuise, M. and Zalcberg, J.R. (1999) Altered expression of the IGF-1 receptor in a tamoxifen-resistant human breast cancer cell line. *British journal of cancer*, **79**, 693-700.
39. Hutcheson, I.R., Knowlden, J.M., Madden, T.A., Barrow, D., Gee, J.M., Wakeling, A.E. and Nicholson, R.I. (2003) Oestrogen receptor-mediated modulation of the EGFR/MAPK pathway in tamoxifen-resistant MCF-7 cells. *Breast cancer research and treatment*, **81**, 81-93.
40. Contreras, A., Herrera, S., Wang, T., Mayer, I., Forero, A., Nanda, R., Goetz, M., Chang, J., Pavlick, A., Fuqua, S. *et al.* (2013) Abstract PD1-2: PIK3CA mutations and/or low PTEN predict resistance to combined anti-HER2 therapy with lapatinib and trastuzumab and without chemotherapy in TBCRC006, a neoadjuvant trial of HER2-positive breast cancer patients. *Cancer research*, **73**, PD1-2.
41. Kandioler-Eckersberger, D., Ludwig, C., Rudas, M., Kappel, S., Janschek, E., Wenzel, C., Schlagbauer-Wadl, H., Mittlbock, M., Gnant, M., Steger, G. *et al.* (2000) TP53 mutation and p53 overexpression for prediction of response to neoadjuvant treatment in breast cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **6**, 50-56.
42. Schroth, W., Goetz, M.P., Hamann, U., Fasching, P.A., Schmidt, M., Winter, S., Fritz, P., Simon, W., Suman, V.J., Ames, M.M. *et al.* (2009) Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. *Jama*, **302**, 1429-1436.
43. Wegman, P., Elingarami, S., Carstensen, J., Stal, O., Nordenskjold, B. and Wingren, S. (2007) Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer. *Breast cancer research : BCR*, **9**, R7.
44. Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D. and Carroll, J.S. (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature genetics*, **43**, 27-33.
45. Ross-Innes, C.S., Stark, R., Holmes, K.A., Schmidt, D., Spyrou, C., Russell, R., Massie, C.E., Vowler, S.L., Eldridge, M. and Carroll, J.S. (2010) Cooperative interaction between retinoic acid receptor-alpha and estrogen receptor in breast cancer. *Genes & development*, **24**, 171-182.
46. Tsai, W.W., Wang, Z., Yiu, T.T., Akdemir, K.C., Xia, W., Winter, S., Tsai, C.Y., Shi, X., Schwarzer, D., Plunkett, W. *et al.* (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature*, **468**, 927-932.
47. Yu, J.C., Hsiung, C.N., Hsu, H.M., Bao, B.Y., Chen, S.T., Hsu, G.C., Chou, W.C., Hu, L.Y., Ding, S.L., Cheng, C.W. *et al.* (2011) Genetic variation in the genome-wide predicted estrogen response element-related sequences is associated with breast cancer development. *Breast cancer research : BCR*, **13**, R13.
48. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struwing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087-1093.

49. Meyer, K.B., O'Reilly, M., Michailidou, K., Carlebur, S., Edwards, S.L., French, J.D., Prathalingham, R., Dennis, J., Bolla, M.K., Wang, Q. *et al.* (2013) Fine-scale mapping of the FGFR2 breast cancer risk locus: putative functional variants differentially bind FOXA1 and E2F1. *American journal of human genetics*, **93**, 1046-1060.
50. Cowper-Salari, R., Zhang, X., Wright, J.B., Bailey, S.D., Cole, M.D., Eeckhoutte, J., Moore, J.H. and Lupien, M. (2012) Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. *Nature genetics*, **44**, 1191-1198.
51. Ingle, J.N., Liu, M., Wickerham, D.L., Schaid, D.J., Wang, L., Mushiroda, T., Kubo, M., Costantino, J.P., Vogel, V.G., Paik, S. *et al.* (2013) Selective estrogen receptor modulators and pharmacogenomic variation in ZNF423 regulation of BRCA1 expression: individualized breast cancer prevention. *Cancer discovery*, **3**, 812-825.
52. Ingle, J.N., Schaid, D.J., Goss, P.E., Liu, M., Mushiroda, T., Chapman, J.A., Kubo, M., Jenkins, G.D., Batzler, A., Shepherd, L. *et al.* (2010) Genome-wide associations and functional genomic studies of musculoskeletal adverse events in women receiving aromatase inhibitors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **28**, 4674-4682.
53. Garcia, T., Sanchez, M., Cox, J.L., Shaw, P.A., Ross, J.B., Lehrer, S. and Schachter, B. (1989) Identification of a variant form of the human estrogen receptor with an amino acid replacement. *Nucleic acids research*, **17**, 8364.
54. Karnik, P.S., Kulkarni, S., Liu, X.P., Budd, G.T. and Bukowski, R.M. (1994) Estrogen receptor mutations in tamoxifen-resistant breast cancer. *Cancer research*, **54**, 349-353.
55. McGuire, W.L., Chamness, G.C. and Fuqua, S.A. (1992) Abnormal estrogen receptor in clinical breast cancer. *The Journal of steroid biochemistry and molecular biology*, **43**, 243-247.
56. Roodi, N., Bailey, L.R., Kao, W.Y., Verrier, C.S., Yee, C.J., Dupont, W.D. and Parl, F.F. (1995) Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. *Journal of the National Cancer Institute*, **87**, 446-451.
57. Zhang, Q.X., Borg, A., Wolf, D.M., Oesterreich, S. and Fuqua, S.A. (1997) An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer. *Cancer research*, **57**, 1244-1249.
58. Cancer Genome Atlas, N. (2012) Comprehensive molecular portraits of human breast tumours. *Nature*, **490**, 61-70.
59. Li, S., Shen, D., Shao, J., Crowder, R., Liu, W., Prat, A., He, X., Liu, S., Hoog, J., Lu, C. *et al.* (2013) Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell reports*, **4**, 1116-1130.
60. Jeselsohn, R., Yelensky, R., Buchwalter, G., Frampton, G., Meric-Bernstam, F., Gonzalez-Angulo, A.M., Ferrer-Lozano, J., Perez-Fidalgo, J.A., Cristofanilli, M., Gomez, H. *et al.* (2014) Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **20**, 1757-1767.
61. Merenbakh-Lamin, K., Ben-Baruch, N., Yeheskel, A., Dvir, A., Soussan-Gutman, L., Jeselsohn, R., Yelensky, R., Brown, M., Miller, V.A., Sarid, D. *et al.* (2013) D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer. *Cancer research*, **73**, 6856-6864.

62. Robinson, D.R., Wu, Y.M., Vats, P., Su, F., Lonigro, R.J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L. *et al.* (2013) Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nature genetics*, **45**, 1446-1451.
63. Toy, W., Shen, Y., Won, H., Green, B., Sakr, R.A., Will, M., Li, Z., Gala, K., Fanning, S., King, T.A. *et al.* (2013) ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nature genetics*, **45**, 1439-1445.
64. Sefrioui, D., Perdrix, A., Sarafan-Vasseur, N., Dolfus, C., Dujon, A., Picquenot, J.M., Delacour, J., Cornic, M., Bohers, E., Leheurteur, M. *et al.* (2015) Short report: Monitoring ESR1 mutations by circulating tumor DNA in aromatase inhibitor resistant metastatic breast cancer. *International journal of cancer. Journal international du cancer*, **137**, 2513-2519.
65. Schiavon, G., Hrebien, S., Garcia-Murillas, I., Cutts, R.J., Pearson, A., Tarazona, N., Fenwick, K., Kozarewa, I., Lopez-Knowles, E., Ribas, R. *et al.* (2015) Analysis of ESR1 mutation in circulating tumor DNA demonstrates evolution during therapy for metastatic breast cancer. *Science translational medicine*, **7**, 313ra182.
66. Chu, D., Paoletti, C., Gersch, C., VanDenBerg, D.A., Zabransky, D.J., Cochran, R.L., Wong, H.Y., Toro, P.V., Cidado, J., Croessmann, S. *et al.* (2015) ESR1 Mutations in Circulating Plasma Tumor DNA from Metastatic Breast Cancer Patients. *Clinical cancer research : an official journal of the American Association for Cancer Research*.
67. Wang, P., Bahreini, A., Gyanchandani, R., Lucas, P.C., Hartmaier, R.J., Watters, R.J., Jonnalagadda, A.R., Trejo Bittar, H.E., Berg, A., Hamilton, R.L. *et al.* (2015) Sensitive detection of mono- and polyclonal ESR1 mutations in primary tumors, metastatic lesions and cell free DNA of breast cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research*.
68. De Mattos-Arruda, L., Weigelt, B., Cortes, J., Won, H.H., Ng, C.K., Nuciforo, P., Bidard, F.C., Aura, C., Saura, C., Peg, V. *et al.* (2014) Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*, **25**, 1729-1735.
69. Yu, M., Bardia, A., Aceto, N., Bersani, F., Madden, M.W., Donaldson, M.C., Desai, R., Zhu, H., Comaills, V., Zheng, Z. *et al.* (2014) Cancer therapy. Ex vivo culture of circulating breast tumor cells for individualized testing of drug susceptibility. *Science*, **345**, 216-220.
70. Carlson, K.E., Choi, I., Gee, A., Katzenellenbogen, B.S. and Katzenellenbogen, J.A. (1997) Altered ligand binding properties and enhanced stability of a constitutively active estrogen receptor: evidence that an open pocket conformation is required for ligand interaction. *Biochemistry*, **36**, 14897-14905.
71. Weis, K.E., Ekena, K., Thomas, J.A., Lazennec, G. and Katzenellenbogen, B.S. (1996) Constitutively active human estrogen receptors containing amino acid substitutions for tyrosine 537 in the receptor protein. *Molecular endocrinology*, **10**, 1388-1398.
72. Feng, J., Yan, J., Michaud, S., Craddock, N., Jones, I.R., Cook, E.H., Jr., Goldman, D., Heston, L.L., Peltonen, L., Delisi, L.E. *et al.* (2001) Scanning of estrogen receptor alpha (ERalpha) and thyroid hormone receptor alpha (TRalpha) genes in patients with psychiatric diseases: four missense mutations identified in ERalpha gene. *American journal of medical genetics*, **105**, 369-374.



73. Zelada-Hedman, M., Borresen-Dale, A.L. and Lindblom, A. (1997) Screening of 229 family cancer patients for a germline estrogen receptor gene (ESR) base mutation. *Human mutation*, **9**, 289.
74. Fuqua, S.A., Wiltschke, C., Zhang, Q.X., Borg, A., Castles, C.G., Friedrichs, W.E., Hopp, T., Hilsenbeck, S., Mohsin, S., O'Connell, P. *et al.* (2000) A hypersensitive estrogen receptor-alpha mutation in premalignant breast lesions. *Cancer research*, **60**, 4026-4029.
75. Siegel, R., Ma, J., Zou, Z. and Jemal, A. (2014) Cancer statistics, 2014. *CA: a cancer journal for clinicians*, **64**, 9-29.
76. Smith, R.A., Cokkinides, V. and Eyre, H.J. (2006) American Cancer Society guidelines for the early detection of cancer, 2006. *CA: a cancer journal for clinicians*, **56**, 11-25; quiz 49-50.
77. Couch, F.J., DeShano, M.L., Blackwood, M.A., Calzone, K., Stopfer, J., Campeau, L., Ganguly, A., Rebbeck, T. and Weber, B.L. (1997) BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. *The New England journal of medicine*, **336**, 1409-1415.
78. Narod, S.A. and Offit, K. (2005) Prevention and management of hereditary breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, **23**, 1656-1663.
79. Olopade, O.I., Grushko, T.A., Nanda, R. and Huo, D. (2008) Advances in breast cancer: pathways to personalized medicine. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **14**, 7988-7999.
80. Consortium, E.P. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636-640.
81. Weinhold, N., Jacobsen, A., Schultz, N., Sander, C. and Lee, W. (2014) Genome-wide analysis of noncoding regulatory mutations in cancer. *Nature genetics*, **46**, 1160-1165.
82. Fredriksson, N.J., Ny, L., Nilsson, J.A. and Larsson, E. (2014) Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nature genetics*, **46**, 1258-1263.
83. Park, P.J. (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nature reviews. Genetics*, **10**, 669-680.
84. Ameer, A., Rada-Iglesias, A., Komorowski, J. and Wadelius, C. (2009) Identification of candidate regulatory SNPs by combination of transcription-factor-binding site prediction, SNP genotyping and haploChIP. *Nucleic acids research*, **37**, e85.
85. Bryzgalov, L.O., Antontseva, E.V., Matveeva, M.Y., Shilov, A.G., Kashina, E.V., Mordvinov, V.A. and Merkulova, T.I. (2013) Detection of regulatory SNPs in human genome using ChIP-seq ENCODE data. *PloS one*, **8**, e78833.
86. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297-1303.
87. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
88. Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J., Wallis, J.W., Van Tine, B.A., Hoog, J., Goiffon, R.J., Goldstein, T.C. *et al.* (2012) Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, **486**, 353-360.

89. Larson, D.E., Harris, C.C., Chen, K., Koboldt, D.C., Abbott, T.E., Dooling, D.J., Ley, T.J., Mardis, E.R., Wilson, R.K. and Ding, L. (2012) SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*, **28**, 311-317.
90. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology*, **9**, R137.
91. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. and Lenhard, B. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic acids research*, **32**, D91-94.
92. Macintyre, G., Bailey, J., Haviv, I. and Kowalczyk, A. (2010) is-rSNP: a novel technique for in silico regulatory SNP detection. *Bioinformatics*, **26**, i524-530.
93. Sikora, M.J., Cooper, K.L., Bahreini, A., Luthra, S., Wang, G., Chandran, U.R., Davidson, N.E., Dabbs, D.J., Welm, A.L. and Oesterreich, S. (2014) Invasive lobular carcinoma cell lines are characterized by unique estrogen-mediated gene expression patterns and altered tamoxifen response. *Cancer research*, **74**, 1463-1474.
94. Guan, Y., Kuo, W.L., Stilwell, J.L., Takano, H., Lapuk, A.V., Fridlyand, J., Mao, J.H., Yu, M., Miller, M.A., Santos, J.L. *et al.* (2007) Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **13**, 5745-5755.
95. Werner, H. and Bruchim, I. (2009) The insulin-like growth factor-I receptor as an oncogene. *Archives of physiology and biochemistry*, **115**, 58-71.
96. Rae, J.M., Johnson, M.D., Scheys, J.O., Cordero, K.E., Larios, J.M. and Lippman, M.E. (2005) GREB 1 is a critical regulator of hormone dependent breast cancer growth. *Breast cancer research and treatment*, **92**, 141-149.
97. Welboren, W.J., van Driel, M.A., Janssen-Megens, E.M., van Heeringen, S.J., Sweep, F.C., Span, P.N. and Stunnenberg, H.G. (2009) ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *The EMBO journal*, **28**, 1418-1428.
98. Sheehan, D., Meade, G., Foley, V.M. and Dowd, C.A. (2001) Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *The Biochemical journal*, **360**, 1-16.
99. Bell, D.A., Taylor, J.A., Paulson, D.F., Robertson, C.N., Mohler, J.L. and Lucier, G.W. (1993) Genetic risk and carcinogen exposure: a common inherited defect of the carcinogen-metabolism gene glutathione S-transferase M1 (GSTM1) that increases susceptibility to bladder cancer. *Journal of the National Cancer Institute*, **85**, 1159-1164.
100. Garcia, T., Lehrer, S., Bloomer, W.D. and Schachter, B. (1988) A variant estrogen receptor messenger ribonucleic acid is associated with reduced levels of estrogen binding in human mammary tumors. *Molecular endocrinology*, **2**, 785-791.
101. McWilliams, J.E., Sanderson, B.J., Harris, E.L., Richert-Boe, K.E. and Henner, W.D. (1995) Glutathione S-transferase M1 (GSTM1) deficiency and lung cancer risk. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, **4**, 589-594.
102. Spivakov, M., Akhtar, J., Kheradpour, P., Beal, K., Girardot, C., Koscielny, G., Herrero, J., Kellis, M., Furlong, E.E. and Birney, E. (2012) Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome biology*, **13**, R49.

103. Kasowski, M., Grubert, F., Heffelfinger, C., Hariharan, M., Asabere, A., Waszak, S.M., Habegger, L., Rozowsky, J., Shi, M., Urban, A.E. *et al.* (2010) Variation in transcription factor binding among humans. *Science*, **328**, 232-235.
104. Zheng, W., Zhao, H., Mancera, E., Steinmetz, L.M. and Snyder, M. (2010) Genetic analysis of variation in transcription factor binding in yeast. *Nature*, **464**, 1187-1191.
105. Melton, C., Reuter, J.A., Spacek, D.V. and Snyder, M. (2015) Recurrent somatic mutations in regulatory regions of human cancer genomes. *Nature genetics*, **47**, 710-716.
106. Klinakis, A., Szabolcs, M., Chen, G., Xuan, S., Hibshoosh, H. and Efstratiadis, A. (2009) Igf1r as a therapeutic target in a mouse model of basal-like breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 2359-2364.
107. Sachdev, D. and Yee, D. (2006) Inhibitors of insulin-like growth factor signaling: a therapeutic approach for breast cancer. *Journal of mammary gland biology and neoplasia*, **11**, 27-39.
108. Yerushalmi, R., Gelmon, K.A., Leung, S., Gao, D., Cheang, M., Pollak, M., Turashvili, G., Gilks, B.C. and Kennecke, H. (2012) Insulin-like growth factor receptor (IGF-1R) in breast cancer subtypes. *Breast cancer research and treatment*, **132**, 131-142.
109. Fagan, D.H. and Yee, D. (2008) Crosstalk between IGF1R and estrogen receptor signaling in breast cancer. *Journal of mammary gland biology and neoplasia*, **13**, 423-429.
110. Kahlert, S., Nuedling, S., van Eickels, M., Vetter, H., Meyer, R. and Grohe, C. (2000) Estrogen receptor alpha rapidly activates the IGF-1 receptor pathway. *The Journal of biological chemistry*, **275**, 18447-18453.
111. Bojesen, S.E., Pooley, K.A., Johnatty, S.E., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.E. *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nature genetics*, **45**, 371-384, 384e371-372.
112. Glenn, T.D. and Talbot, W.S. (2013) Analysis of Gpr126 function defines distinct mechanisms controlling the initiation and maturation of myelin. *Development*, **140**, 3167-3175.
113. Monk, K.R., Oshima, K., Jors, S., Heller, S. and Talbot, W.S. (2011) Gpr126 is essential for peripheral nerve development and myelination in mammals. *Development*, **138**, 2673-2680.
114. Chimge, N.O., Baniwal, S.K., Little, G.H., Chen, Y.B., Kahn, M., Tripathy, D., Borok, Z. and Frenkel, B. (2011) Regulation of breast cancer metastasis by Runx2 and estrogen signaling: the role of SNAI2. *Breast cancer research : BCR*, **13**, R127.
115. Yager, J.D. and Davidson, N.E. (2006) Estrogen carcinogenesis in breast cancer. *N Engl J Med*, **354**, 270-282.
116. Martz, C.A., Ottina, K.A., Singleton, K.R., Jasper, J.S., Wardell, S.E., Peraza-Penton, A., Anderson, G.R., Winter, P.S., Wang, T., Alley, H.M. *et al.* (2014) Systematic identification of signaling pathways with potential to confer anticancer drug resistance. *Sci Signal*, **7**, ra121.
117. Veeraraghavan, J., Tan, Y., Cao, X.X., Kim, J.A., Wang, X., Chamness, G.C., Maiti, S.N., Cooper, L.J., Edwards, D.P., Contreras, A. *et al.* (2014) Recurrent ESR1-CCDC170 rearrangements in an aggressive subset of oestrogen receptor-positive breast cancers. *Nat Commun*, **5**, 4577.

118. Rothe, F., Laes, J.F., Lambrechts, D., Smeets, D., Vincent, D., Maetens, M., Fumagalli, D., Michiels, S., Drisis, S., Moerman, C. *et al.* (2014) Plasma circulating tumor DNA as an alternative to metastatic biopsies for mutational analysis in breast cancer. *Ann Oncol*, **25**, 1959-1965.
119. Bettgowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Lubner, B., Alani, R.M. *et al.* (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*, **6**, 224ra224.
120. Dawson, S.J., Tsui, D.W., Murtaza, M., Biggs, H., Rueda, O.M., Chin, S.F., Dunning, M.J., Gale, D., Forshew, T., Mahler-Araujo, B. *et al.* (2013) Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med*, **368**, 1199-1209.
121. Beaver, J.A., Jelovac, D., Balukrishna, S., Cochran, R.L., Croessmann, S., Zabransky, D.J., Wong, H.Y., Valda Toro, P., Cidado, J., Blair, B.G. *et al.* (2014) Detection of cancer DNA in plasma of patients with early-stage breast cancer. *Clin Cancer Res*, **20**, 2643-2650.
122. Day, E., Dear, P.H. and McCaughan, F. (2013) Digital PCR strategies in the development and analysis of molecular biomarkers for personalized medicine. *Methods*, **59**, 101-107.
123. Hammond, M.E., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M. *et al.* (2010) American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol*, **28**, 2784-2795.
124. Conway, K., Parrish, E., Edmiston, S.N., Tolbert, D., Tse, C.K., Geradts, J., Livasy, C.A., Singh, H., Newman, B. and Millikan, R.C. (2005) The estrogen receptor-alpha A908G (K303R) mutation occurs at a low frequency in invasive breast tumors: results from a population-based study. *Breast Cancer Res*, **7**, R871-880.
125. Davies, M.P., O'Neill, P.A., Innes, H. and Sibson, D.R. (2005) Hypersensitive K303R oestrogen receptor-alpha variant not found in invasive carcinomas. *Breast Cancer Res*, **7**, R113-118.
126. Ghimenti, C., Mello-Grand, M., Regolo, L., Zambelli, A. and Chiorino, G. (2010) Absence of the K303R estrogen receptor alpha mutation in breast cancer patients exhibiting different responses to aromatase inhibitor anastrozole neoadjuvant treatment. *Exp Ther Med*, **1**, 939-942.
127. Tokunaga, E., Kimura, Y. and Maehara, Y. (2004) No hypersensitive estrogen receptor-alpha mutation (K303R) in Japanese breast carcinomas. *Breast Cancer Res Treat*, **84**, 289-292.
128. Herynk, M.H., Parra, I., Cui, Y., Beyer, A., Wu, M.F., Hilsenbeck, S.G. and Fuqua, S.A. (2007) Association between the estrogen receptor alpha A908G mutation and outcomes in invasive breast cancer. *Clin Cancer Res*, **13**, 3235-3243.
129. Abbasi, S., Rasouli, M., Nouri, M. and Kalbasi, S. (2013) Association of estrogen receptor-alpha A908G (K303R) mutation with breast cancer risk. *Int J Clin Exp Med*, **6**, 39-49.
130. Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P. *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med*, **366**, 883-892.

131. Chu, D., Paoletti, C., Gersch, C., VanDenBerg, D., Zabransky, D., Cochran, R., Wong, H.Y., Valda Toro, P., Cidado, J., Croessmann, S. *et al.* (2015) ESR1 mutations in circulating plasma tumor DNA from metastatic breast cancer patients. *Clin Cancer Res.*
132. Garcia-Murillas, I., Schiavon, G., Weigelt, B., Ng, C., Hrebien, S., Cutts, R.J., Cheang, M., Osin, P., Nerurkar, A., Kozarewa, I. *et al.* (2015) Mutation tracking in circulating tumor DNA predicts relapse in early breast cancer. *Sci Transl Med*, **7**, 302ra133.
133. Lannigan, D.A. (2003) Estrogen receptor phosphorylation. *Steroids*, **68**, 1-9.
134. Kumar, V. and Jain, M. (2015) The CRISPR-Cas system for plant genome editing: advances and opportunities. *Journal of experimental botany*, **66**, 47-57.
135. Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F. and Jaenisch, R. (2013) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, **153**, 910-918.
136. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173-1183.
137. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.
138. Zabransky, D.J., Yankaskas, C.L., Cochran, R.L., Wong, H.Y., Croessmann, S., Chu, D., Kavuri, S.M., Red Brewer, M., Rosen, D.M., Dalton, W.B. *et al.* (2015) HER2 missense mutations have distinct effects on oncogenic signaling and migration. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, E6205-6214.
139. Patro, R., Duggal, G. and Kingsford, C. (2015) Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment. *bioRxiv*.
140. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 550.
141. Kato, S., Endoh, H., Masuhiro, Y., Kitamoto, T., Uchiyama, S., Sasaki, H., Masushige, S., Gotoh, Y., Nishida, E., Kawashima, H. *et al.* (1995) Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science*, **270**, 1491-1494.
142. Clevenger, C.V. (2004) Roles and regulation of stat family transcription factors in human breast cancer. *The American journal of pathology*, **165**, 1449-1460.
143. Doherty, G.M., Boucher, L., Sorenson, K. and Lowney, J. (2001) Interferon regulatory factor expression in human breast cancer. *Annals of surgery*, **233**, 623-629.
144. van den Berg, H.W., Leahey, W.J., Lynch, M., Clarke, R. and Nelson, J. (1987) Recombinant human interferon alpha increases oestrogen receptor expression in human breast cancer cells (ZR-75-1) and sensitizes them to the anti-proliferative effects of tamoxifen. *British journal of cancer*, **55**, 255-257.
145. Buck, M.B. and Knabbe, C. (2006) TGF-beta signaling in breast cancer. *Annals of the New York Academy of Sciences*, **1089**, 119-126.
146. Deckers, M., van Dinther, M., Buijs, J., Que, I., Lowik, C., van der Pluijm, G. and ten Dijke, P. (2006) The tumor suppressor Smad4 is required for transforming growth factor beta-induced epithelial to mesenchymal transition and bone metastasis of breast cancer cells. *Cancer research*, **66**, 2202-2209.
147. Dumont, N. and Arteaga, C.L. (2000) Transforming growth factor-beta and breast cancer: Tumor promoting effects of transforming growth factor-beta. *Breast cancer research : BCR*, **2**, 125-132.

148. Petz, L.N., Ziegler, Y.S., Schultz, J.R. and Nardulli, A.M. (2004) Fos and Jun inhibit estrogen-induced transcription of the human progesterone receptor gene through an activator protein-1 site. *Molecular endocrinology*, **18**, 521-532.
149. Piek, E., Moustakas, A., Kurisaki, A., Heldin, C.H. and ten Dijke, P. (1999) TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *Journal of cell science*, **112** ( Pt **24**), 4557-4568.
150. Stuelten, C.H., Buck, M.B., Dippon, J., Roberts, A.B., Fritz, P. and Knabbe, C. (2006) Smad4-expression is decreased in breast cancer tissues: a retrospective study. *BMC cancer*, **6**, 25.
151. Wagstaff, S.C., Bowler, W.B., Gallagher, J.A. and Hipskind, R.A. (2000) Extracellular ATP activates multiple signalling pathways and potentiates growth factor-induced c-fos gene expression in MCF-7 breast cancer cells. *Carcinogenesis*, **21**, 2175-2181.
152. Wiercinska, E., Naber, H.P., Pardali, E., van der Pluijm, G., van Dam, H. and ten Dijke, P. (2011) The TGF-beta/Smad pathway induces breast cancer cell invasion through the up-regulation of matrix metalloproteinase 2 and 9 in a spheroid invasion model system. *Breast cancer research and treatment*, **128**, 657-666.
153. Behrens, J. (1993) The role of cell adhesion molecules in cancer invasion and metastasis. *Breast cancer research and treatment*, **24**, 175-184.
154. Kleinman, H.K., Klebe, R.J. and Martin, G.R. (1981) Role of collagenous matrices in the adhesion and growth of cells. *The Journal of cell biology*, **88**, 473-485.
155. Herbst, T.J., McCarthy, J.B., Tsilibary, E.C. and Furcht, L.T. (1988) Differential effects of laminin, intact type IV collagen, and specific domains of type IV collagen on endothelial cell adhesion and migration. *The Journal of cell biology*, **106**, 1365-1373.
156. Liotta, L.A., Tryggvason, K., Garbisa, S., Hart, I., Foltz, C.M. and Shafie, S. (1980) Metastatic potential correlates with enzymatic degradation of basement membrane collagen. *Nature*, **284**, 67-68.
157. Shen, Y., Shen, R., Ge, L., Zhu, Q. and Li, F. (2012) Fibrillar type I collagen matrices enhance metastasis/invasion of ovarian epithelial cancer via beta1 integrin and PTEN signals. *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, **22**, 1316-1324.
158. Shintani, Y., Hollingsworth, M.A., Wheelock, M.J. and Johnson, K.R. (2006) Collagen I promotes metastasis in pancreatic cancer by activating c-Jun NH(2)-terminal kinase 1 and up-regulating N-cadherin expression. *Cancer research*, **66**, 11745-11753.
159. Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O. *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, **31**, 827-832.
160. Sarwar, N., Kim, J.S., Jiang, J., Peston, D., Sinnett, H.D., Madden, P., Gee, J.M., Nicholson, R.I., Lykkesfeldt, A.E., Shousha, S. *et al.* (2006) Phosphorylation of ERalpha at serine 118 in primary breast cancer and in tamoxifen-resistant tumours is indicative of a complex role for ERalpha phosphorylation in breast cancer progression. *Endocrine-related cancer*, **13**, 851-861.
161. Honn, K.V. and Tang, D.G. (1992) Adhesion molecules and tumor cell interaction with endothelium and subendothelial matrix. *Cancer metastasis reviews*, **11**, 353-375.
162. Tang, D.G. and Honn, K.V. (1994) Adhesion molecules and tumor metastasis: an update. *Invasion & metastasis*, **14**, 109-122.

163. Zetter, B.R. (1993) Adhesion molecules in tumor metastasis. *Seminars in cancer biology*, **4**, 219-229.
164. Haier, J., Nasralla, M. and Nicolson, G.L. (1999) Different adhesion properties of highly and poorly metastatic HT-29 colon carcinoma cells with extracellular matrix components: role of integrin expression and cytoskeletal components. *British journal of cancer*, **80**, 1867-1874.
165. Satoh, S., Hinoda, Y., Hayashi, T., Burdick, M.D., Imai, K. and Hollingsworth, M.A. (2000) Enhancement of metastatic properties of pancreatic cancer cells by MUC1 gene encoding an anti-adhesion molecule. *International journal of cancer. Journal international du cancer*, **88**, 507-518.